Multi-Microphone Speaker Localization on Manifolds
Achievements and Challenges

Sharon Gannot
joint work with Bracha Laufer-Goldshtein and Ronen Talmon

Faculty of Engineering, Bar-Ilan University, Israel

LVA/ICA, February 21st, 2017
Speech Processing

Noisy and reverberant

1. Train stations
2. Factories
3. Cars
4. Home
5. Busy offices
6. Cocktail parties

Common applications

1. Hands-free communications
2. Teleconferencing (Skype)
3. Assisting hearing impaired
4. Robust automatic speech recognition (ASR)
5. Eavesdropping

Speech enhancement tasks

1. Noise reduction
2. Speaker segregation, separation and extraction
3. Dereverberation
4. Echo cancellation
Speaker Localization

Why localizing?

1. An essential component in many speech enhancement algorithms
2. Camera steering
3. Robot audition
4. Simultaneous localization and mapping (SLAM)

1. Accuracy deteriorates in presence of noise
2. Reverberation severely degrades accuracy if not taken into account (as in many speech processing tasks)
3. Necessitates multi-microphone installations (single mic. [Talmon et al., 2011])
Devices Equipped with Multiple Microphones

1. Cellular phones
2. Laptops and tablets
3. Hearing devices
4. Smart watches
5. Smart glasses
6. Smart homes & cars
Room Acoustics Essentials

Acoustic propagation models

- When sound propagates in an enclosure it undergoes reflections from the room surfaces.
- Reflections can be modeled as images beyond room walls and hence impinging the microphones from many directions [Allen and Berkley, 1979, Peterson, 1986]
- Late reflections tend to be diffused, hence do not exhibit directionality [Dal-Degan and Prati, 1988, Habets and Gannot, 2007]
Speech Processing in Acoustic Environments

- Classical multi-microphone speech processing algorithms use time difference of arrival (TDOA)-only model.
- Viable speech processing solutions can only be accomplished by an accurate source propagation description, captured by the acoustic impulse response (AIR).
- Describing the wave propagation of any audio source in an arbitrary acoustic environment is, however, a cumbersome task, since:
  - No simple mathematical models exist.
  - The estimation of the vast number of parameters used to describe the wave propagation suffers from large errors.

Data-driven approach

To alleviate these limitations and to infer a mathematical model that is accurate, simple to describe and simple to implement, we propose a data-driven approach.
How to model the Acoustic Environment?

How to model the acoustic path?

- TDOA-only
  - Too simplified
  - Interesting...

- Manifold Learning
  - Interesting...

- Entire Impulse Response
  - Perfect but complicated

S. Gannot (BIU)
Main Claims

- The variability of the acoustic response in specific enclosures depends only on a small number of parameters.
- The intrinsic degrees of freedom in acoustic responses are limited to a small number of variables.

$\Rightarrow$ manifold learning approaches may improve localization ability.

Controlling Parameters

- room dimensions
- reverberation time
- microphone position
- source position
- ...

Controlling Parameters

Acoustic Impulse Response

RT60

0.01
0.02
0.03
0.04
0.05
0
500
1000
1500
Outline

1. Data model and acoustic features
2. The Acoustic Manifold
3. Data-Driven Source Localization: Microphone Pair
4. Bayesian Perspective
5. Data-Driven Source Localization: Ad Hoc Array
Outline

1. Data model and acoustic features
2. The Acoustic Manifold
3. Data-Driven Source Localization: Microphone Pair
4. Bayesian Perspective
5. Data-Driven Source Localization: Ad Hoc Array
Data Model: The Two Microphone Case

Microphone signals:

The measured signals in the two microphones:

\[ y_1(n) = a_1(n) \ast s(n) + u_1(n) \]
\[ y_2(n) = a_2(n) \ast s(n) + u_2(n) \]

- \( s(n) \) - the source signal
- \( a_i(n), \ i = \{1, 2\} \) - the acoustic impulse responses relating the source and each of the microphones
- \( u_i(n), \ i = \{1, 2\} \) - noise signals, independent of the source
Data Model: The Two Microphone Case

Microphone signals:
The measured signals in the two microphones:

\[ y_1(n) = a_1(n) \ast s(n) + u_1(n) \]
\[ y_2(n) = a_2(n) \ast s(n) + u_2(n) \]

- \( s(n) \) - the source signal
- \( a_i(n), i = \{1, 2\} \) - the acoustic impulse responses relating the source and each of the microphones
- \( u_i(n), i = \{1, 2\} \) - noise signals, independent of the source

Find a feature vector representing the characteristics of the acoustic path and independent of the source signal!
Relative Transfer Function (RTF) [Gannot et al., 2001]

- Defined as the ratio between the transfer functions of the two mics:

\[
H_{12}(k) = \frac{A_2(k)}{A_1(k)}
\]

- In the time domain: the relative impulse response (RIR) satisfies:

\[
a_2(n) = h_{12}(n) \ast a_1(n)
\]
Relative Transfer Function (RTF) [Gannot et al., 2001]

\[
\frac{\text{RTF}}{\text{RTF}} = f \left( \begin{array}{c}
\text{room dimensions} \\
\text{reverberation time} \\
\text{microphone position} \\
\text{source position} \\
\text{...}
\end{array} \right)
\]

**RTF:**
- Represents the acoustic paths and is independent of the source signal
- Generalizes the TDOA
- Depends on a small set of parameters related to the physical characteristics of the environment
Relative Transfer Function (RTF) \cite{Gannot et al., 2001}

- room dimensions
- reverberation time
- microphone position
- source position
- ...

**RTF:**

- Represents the acoustic paths and is independent of the source signal
- Generalizes the TDOA
- Depends on a small set of parameters related to the physical characteristics of the environment
- In a static environment the source position is the only varying degree of freedom
Relative Transfer Function (RTF)

RTF-based feature vector:

- Estimated based on PSD and cross-PSD
  (alternatively [Markovich-Golan and Gannot, 2015, Koldovsky et al., 2014]):

\[
\hat{H}_{12}(k) = \frac{\hat{S}_{y_2y_1}(k)}{\hat{S}_{y_1y_1}(k)} \sim \frac{A_2(k)}{A_1(k)}
\]

- Define the feature vector:

\[
h = \left[ \hat{H}_{12}(k_1), \ldots, \hat{H}_{12}(k_D) \right]^T
\]

- \(D \propto\) length of the relative impulse response (time domain)

High dimensional representation - due to reverberation

Controlled by one dominant factor - source position
<table>
<thead>
<tr>
<th></th>
<th>Outline</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data model and acoustic features</td>
</tr>
<tr>
<td>2</td>
<td>The Acoustic Manifold</td>
</tr>
<tr>
<td>3</td>
<td>Data-Driven Source Localization: Microphone Pair</td>
</tr>
<tr>
<td>4</td>
<td>Bayesian Perspective</td>
</tr>
<tr>
<td>5</td>
<td>Data-Driven Source Localization: Ad Hoc Array</td>
</tr>
</tbody>
</table>
How to Measure the **Affinity** between RTFs? [Laufer-Goldshtein et al., 2015, Laufer-Goldshtein et al., 2016b]

- The RTFs are represented as points in a **high dimensional space**
- Small Euclidean distance of high dimensional vectors implies proximity
- Large Euclidean distance of high dimensional vectors is meaningless
How to Measure the **Affinity** between RTFs? [Laufer-Goldshtein et al., 2015, Laufer-Goldshtein et al., 2016b]

- The RTFs are represented as points in a **high dimensional space**
- Small Euclidean distance of high dimensional vectors implies proximity
- Large Euclidean distance of high dimensional vectors is meaningless

**Acoustic manifold**

- They lie on a **low dimensional nonlinear manifold** $\mathcal{M}$. 
How to Measure the Affinity between RTFs? [Laufer-Goldshtein et al., 2015, Laufer-Goldshtein et al., 2016b]

- The RTFs are represented as points in a high dimensional space.
- Small Euclidean distance of high dimensional vectors implies proximity.
- Large Euclidean distance of high dimensional vectors is meaningless.

Acoustic manifold

- They lie on a low dimensional nonlinear manifold $\mathcal{M}$.
- Linearity is preserved in small neighbourhoods.
The RTFs are represented as points in a high dimensional space.

- Small Euclidean distance of high dimensional vectors implies proximity.
- Large Euclidean distance of high dimensional vectors is meaningless.

**Acoustic manifold**

- They lie on a low dimensional nonlinear manifold $\mathcal{M}$.
- Linearity is preserved in small neighbourhoods.
- Distances between RTFs should be measured along the manifold.
Each distance measure relies on a different *hidden assumption* about the underlying structure of the RTF samples.
Euclidean Distance

The Euclidean distance between RTFs:

$$D_{\text{Euc}}(h_i, h_j) = \|h_i - h_j\|$$

- Compares two RTFs in their original space
- Does not assume an existence of a manifold
- Respects flat manifolds

A good affinity measure only when the RTFs are uniformly scattered all over the space, or when they lie on a flat manifold.
PCA-Based Distance

PCA algorithm

- **The principal components** - the $d$ dominant eigenvectors $\{v_i\}_{i=1}^d$ of the covariance matrix of the data
- The RTFs are **linearly projected** onto the principal components:
  \[
  \nu(h_i) = [v_1, \ldots v_d]^T (h_i - \mu).
  \]

PCA-based distance between RTFs

\[
D_{PCA}(h_i, h_j) = \|\nu(h_i) - \nu(h_j)\|.
\]

- A **global approach** - extracts principal directions of the entire set
- **Linear projections** - the manifold is assumed to be linear/flat
PCA-Based Distance

PCA algorithm

- The principal components - the $d$ dominant eigenvectors $\{v_i\}_{i=1}^d$ of the covariance matrix of the data
- The RTFs are linearly projected onto the principal components:
  \[ \nu(h_i) = [v_1, \ldots, v_d]^T (h_i - \mu). \]

PCA-based distance between RTFs

\[ D_{\text{PCA}}(h_i, h_j) = \|\nu(h_i) - \nu(h_j)\|. \]

- A global approach - extracts principal directions of the entire set
- Linear projections - the manifold is assumed to be linear/flat
PCA-Based Distance

**PCA algorithm**

- **The principal components** - the $d$ dominant eigenvectors $\{v_i\}_{i=1}^d$ of the covariance matrix of the data.
- The RTFs are **linearly projected** onto the principal components:
  \[
  \nu(h_i) = [v_1, \ldots, v_d]^T (h_i - \mu).
  \]

**PCA-based distance between RTFs**

\[
D_{PCA}(h_i, h_j) = \|\nu(h_i) - \nu(h_j)\|.
\]

- **A global approach** - extracts principal directions of the entire set.
- **Linear projections** - the manifold is assumed to be linear/flat.
Diffusion Maps

Discretization of the manifold

- The manifold can be empirically represented by a graph:
  - The RTF samples are the graph nodes.
  - The weights of the edges are defined using a kernel function:

\[
K_{ij} = k(h_i, h_j) = \exp \left\{ -\frac{\|h_i - h_j\|^2}{\varepsilon} \right\}.
\]
The manifold can be empirically represented by a graph:

- The RTF samples are the graph nodes.
- The weights of the edges are defined using a kernel function:

\[ K_{ij} = k(h_i, h_j) = \exp \left\{ -\frac{\|h_i - h_j\|^2}{\varepsilon} \right\}. \]

Define a Markov process on the graph by the transition matrix:

\[ p(h_i, h_j) = P_{ij} = \frac{K_{ij}}{\sum_{r=1}^N K_{ir}}. \]
Diffusion Maps

Discretization of the manifold

- The manifold can be empirically represented by a **graph**:
  - The RTF samples are the **graph nodes**.
  - The weights of the **edges** are defined using a **kernel** function:
    \[ K_{ij} = k(h_i, h_j) = \exp \left\{ -\frac{\|h_i - h_j\|^2}{\varepsilon} \right\}. \]
- Define a **Markov process** on the graph by the **transition matrix**:
  \[ p(h_i, h_j) = P_{ij} = K_{ij} / \sum_{r=1}^{N} K_{ir}. \]

which is a discretization of a **diffusion** process on the manifold.
**Diffusion Maps**

**Diffusion mapping** [Coifman and Lafon, 2006]

- Apply **eigenvalue decomposition (EVD)** to the matrix $P$ and obtain the eigenvalues $\{\lambda_j\}$ and right eigenvectors $\{\varphi_j\}$.
- A **nonlinear mapping** into a new **low-dimensional** Euclidean space:

$$\Phi_d : h_i \mapsto [\lambda_1 \varphi_1^{(i)}, \ldots, \lambda_d \varphi_d^{(i)}]^T.$$  

The mapping provides a **parametrization** of the manifold and represents the **latent variables** - Here, the position of the source.
**Diffusion Distance**

Diffusion distance between RTFs

The distance along the manifold is approximated by the **diffusion distance**:

\[
D_{\text{Diff}}^2(h_i, h_j) = \sum_{r=1}^{N} \frac{(p(h_i, h_r) - p(h_j, h_r))^2}{\phi_0^r}
\]

- Two points are close if they are highly connected in the graph.
- The diffusion distance can be well approximated by the Euclidean distance in the embedded domain:

\[
D_{\text{Diff}}(h_i, h_j) \approx \| \Phi_d(h_i) - \Phi_d(h_j) \|
\]
Diffusion Distance

Illustration
How to measure Affinity?

- Euclidean Distance
- Uniform Scattering
- No processing

RTF samples
How to measure Affinity?

- Euclidean Distance
- PCA-based Distance
- RTF samples
- Uniform Scattering
- Flat Manifold
- No processing
- Global Processing
How to measure Affinity?

- Euclidean Distance
- Diffusion Distance
- PCA-based Distance
- Nonlinear Manifold
- Flat Manifold

RTF samples

No processing

Global Processing

Local & Global Processing

S. Gannot (BIU)
Speaker Localization on Manifolds
LVA/ICA, 21.2.2017
Which of the distance measures is proper?
What is the true underlying structure of the RTFs?
Simulation Results

Room setup

Simulate a reverberant room using the image method [Allen and Berkley, 1979]:

- Room dimension $6 \times 6.2 \times 3\text{m}$
- Microphones at: $[3, 3, 1]$ and $[3.2, 3, 1]$
- The source is positioned at $2\text{m}$ from the mics, the azimuth angle in $10^\circ \div 60^\circ$.
- $T_{60} = 150/300/500\text{ ms}$
- SNR= 20 dB

Test

Measure the distance between each of the RTFs and the RTF corresponding to $10^\circ$:

- If monotonic with respect to the angle - proper distance
- If not monotonic with respect to the angle - improper distance
Euclidean Distance & PCA-based Distance \cite{Laufer-Goldshtein2015}

For both distance measures:
- Monotonic with respect to the angle only in a **limited region**
- This region becomes smaller as the reverberation time increases
- They are inappropriate for measuring angles’ proximity
The diffusion distance:

- Monotonic with respect to the angle for almost the entire range
- It is an appropriate distance measure in terms of the source DOA
- Mapping corresponds well with angles - recovers the latent parameter
Outline

1. Data model and acoustic features
2. The Acoustic Manifold
3. Data-Driven Source Localization: Microphone Pair
4. Bayesian Perspective
5. Data-Driven Source Localization: Ad Hoc Array
Semi-Supervised Learning

Mixed of **supervised** (attached with known locations as anchors) and **unsupervised** (unknown locations) learning
Semi-Supervised Learning

Mixed of **supervised** (attached with known locations as anchors) and **unsupervised** (unknown locations) learning

Why using unlabeled data?

1. **Localization** - training should fit the specific environment of interest
   - Cannot generate a general database for all possible acoustic scenarios
   - Generating a large amount of *labelled data* is cumbersome/impractical
   - Unlabelled data is *freely available* - whenever someone is speaking
Semi-Supervised Learning

Mixed of **supervised** (attached with known locations as anchors) and **unsupervised** (unknown locations) learning

**Why using unlabeled data?**

1. **Localization** - training should fit the specific environment of interest
   - Cannot generate a general database for all possible acoustic scenarios
   - Generating a large amount of **labelled data** is cumbersome/impractical
   - Unlabelled data is **freely available** - whenever someone is speaking

2. Unlabelled data can be utilize to recover the **manifold structure**
Semi-Supervised Learning

Mixed of *supervised* (attached with known locations as anchors) and *unsupervised* (unknown locations) learning

### Why using unlabeled data?

1. **Localization** - training should fit the specific environment of interest
   - Cannot generate a general database for all possible acoustic scenarios
   - Generating a large amount of *labelled data* is cumbersome/impractical
   - Unlabelled data is *freely available* - whenever someone is speaking

2. Unlabelled data can be utilize to recover the *manifold structure*

3. Semi-supervised learning is the natural setting for human learning

---

S. Gannot (BIU)
Semi-Supervised Learning

Unlabelled Samples
Recover the Manifold Structure

Labelled Samples
Anchor Points – Translate RTFs to Positions
Semi-Supervised Learning

Data:
- $H_L = \{h_i\}_{i=1}^{n_L}$ - $n_L$ labelled samples
- $P_L = \{\bar{p}_i\}_{i=1}^{n_L}$ - labels/positions
- $H_U = \{h_i\}_{i=n_L+1}^{n_D}$ - $n_U$ unlabelled samples
- $H_D = H_L \cup H_U$ - $n_D = n_L + n_U$ training samples
- $H_T = \{h_i\}_{i=n_D+1}^{n_T}$ - $n_T$ test samples
Semi-Supervised Learning

**Data:**
- $H_L = \{h_i\}_{i=1}^{n_L}$ - $n_L$ labelled samples
- $P_L = \{\bar{p}_i\}_{i=1}^{n_L}$ - labels/positions
- $H_U = \{h_i\}_{i=n_L+1}^{n_P}$ - $n_U$ unlabelled samples
- $H_D = H_L \cup H_U$ - $n_D = n_L + n_U$ training samples
- $H_T = \{h_i\}_{i=n_D+1}^n$ - $n_T$ test samples

**Goal:** Recover the function $f$ which transforms an RTF to position
Optimization and Manifold Regularization

Optimization in a reproducing kernel Hilbert space (RKHS) \cite{Belkin2006}:

\[
f^* = \arg\min_{f \in \mathcal{H}_k} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(h_i))^2 + \gamma_k \|f\|^2_{\mathcal{H}_k} + \gamma_M \|f\|^2_M
\]
Optimization and Manifold Regularization

Optimization in a reproducing kernel Hilbert space (RKHS) [Belkin et al., 2006]:

\[ f^* = \text{argmin}_{f \in \mathcal{H}_k} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(h_i))^2 + \gamma_k \| f \|^2_{\mathcal{H}_k} + \gamma_M \| f \|^2_M \]

Cost function

\[ \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(h_i))^2 \]

correspondence between function values and labels
Optimization and Manifold Regularization

Optimization in a reproducing kernel Hilbert space (RKHS) [Belkin et al., 2006]:

$$f^* = \arg\min_{f \in \mathcal{H}_k} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(h_i))^2 + \gamma_k \|f\|^2_{\mathcal{H}_k} + \gamma_M \|f\|^2_M$$

- **Cost function**
  $$\frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(h_i))^2$$
- **Tikhonov Regularization**
  $$\|f\|^2_{\mathcal{H}_k}$$
- **correspondence**
  between function values and labels
- **smoothness condition**
  in the RKHS
Optimization in a reproducing kernel Hilbert space (RKHS) [Belkin et al., 2006]:

\[ f^* = \arg\min_{f \in \mathcal{H}_k} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(h_i))^2 + \gamma_k \| f \|^2_{\mathcal{H}_k} + \gamma_M \| f \|^2_{\mathcal{M}} \]
Discretization of the manifold

- The manifold is empirically represented by a graph $G$, with weights:
  \[ W_{ij} = \begin{cases} 
  \exp \left\{ -\frac{\|h_i - h_j\|^2}{\varepsilon_w} \right\} & \text{if } h_j \in \mathcal{N}_i \text{ or } h_i \in \mathcal{N}_j \\
  0 & \text{otherwise}
  \end{cases} \]

  where $\mathcal{N}_j$ is a set consisting of the $d$ nearest-neighbours of $h_j$.

- The graph Laplacian of $G$: $M = S - W$, where $S_{ii} = \sum_{j=1}^{n_D} W_{ij}$.

- Regularization: $\|f\|^2_M = f_D^T M f_D = \frac{1}{2} \sum_{i,j=1}^{n_D} W_{ij} (f(h_i) - f(h_j))^2$

  with $f_D^T = [f_1, f_2, \ldots, f_{n_D}]$
The optimization problem can be recast as:

\[
    f^* = \arg\min_{f \in \mathcal{H}_k} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(h_i))^2 + \gamma_k \|f\|_{\mathcal{H}_k}^2 + \gamma_M f_D^T M f_D
\]
The optimization problem can be recast as:

\[ f^* = \arg\min_{f \in \mathcal{H}_k} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(h_i))^2 + \gamma_k \|f\|_{\mathcal{H}_k}^2 + \gamma_M f_D^T M f_D \]

The representer theorem:

The minimizer over \( \mathcal{H}_k \) of the regularized optimization is represented by:

\[ f^*(h) = \sum_{i=1}^{n_D} a_i k(h_i, h) \]

where \( k : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R} \) is the reproducing kernel of \( \mathcal{H}_k \)

with \( K_{ij} = k(h_i, h_j) = \exp \left\{ -\frac{\|h_i - h_j\|^2}{\varepsilon} \right\} \)
The optimization problem can be recast as:

\[
f^* = \arg \min_{f \in \mathcal{H}_k} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(h_i))^2 + \gamma_k \|f\|_{\mathcal{H}_k}^2 + \gamma M f_D^T M f_D
\]

The representer theorem:

The minimizer over \( \mathcal{H}_k \) of the regularized optimization is represented by:

\[
f^*(h) = \sum_{i=1}^{n_D} a_i k(h_i, h) \quad \Rightarrow \quad \text{closed-form solution for } a^*
\]

where \( k : \mathcal{M} \times \mathcal{M} \to \mathbb{R} \) is the reproducing kernel of \( \mathcal{H}_k \).
Manifold Regularization for Localization (MRL)

[Laufer-Goldshtein et al., 2016c]

System Adaptation

- labelled & unlabelled samples \( \{y_{i1}(n), y_{i2}(n)\}_{i=1}^{n_D} \)
- RTF feature extraction
- \( \{h_i\}_{i=1}^{n_D} \)
- Reproducing kernel & manifold learning
- Interpolation weights
- \( P_L \)
- \( K \)
- \( M \)
- \( \{\hat{a}_i\}_{i=1}^{n_D} \)

Localization

- test sample \( \{y_{t1}(n), y_{t2}(n)\} \)
- RTF feature extraction
- \( h_t \)
- Affinity - new sample vs. training samples via the kernel \( k \)
- \( \{k(h_t, h_j)\}_{i=1}^{n_D} \)
- Position estimation
- \( \hat{f}(h_t) \)

\[ K_{ij} = k(h_i, h_j) \]
Recordings Setup

**Setup:**
- Real recordings carried out at Bar-Ilan acoustic lab
- A $6 \times 6 \times 2.4m$ room controllable reverberation time (set to 620ms)
- Region of interest: a 4m long line at 2.5m distance from the mics
Recordings Setup

Setup:
- Real recordings carried out at Bar-Ilan acoustic lab
- A $6 \times 6 \times 2.4m$ room controllable reverberation time (set to 620ms)
- Region of interest: a 4m long line at 2.5m distance from the mics
Experimental Results [Laufer-Goldshtein et al., 2016c]

Setup:
- Training: 5 labelled samples (1m resolution), 75 unlabelled samples
- Test: 30 random samples in the defined region
- Two noise types: air-conditioner noise and babble noise

Compare with:
- Nearest-neighbour (NN)
- Generalized cross-correlation (GCC) method [Knapp and Carter, 1976]
Experimental Results [Laufer-Goldshtein et al., 2016c]

Setup:
- Training: 5 labelled samples (1m resolution), 75 unlabelled samples
- Test: 30 random samples in the defined region
- Two noise types: air-conditioner noise and babble noise

Compare with:
- Nearest-neighbour (NN)
- Generalized cross-correlation (GCC) method [Knapp and Carter, 1976]

The MRL algorithm outperforms the two other methods.
Effect of Labelled & Unlabelled Samples

<table>
<thead>
<tr>
<th>No. of Labelled Samples</th>
<th>RMSE [m]</th>
<th>NN</th>
<th>MRL</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. of Unabelled Samples</th>
<th>RMSE [m]</th>
<th>MRL (9 labelled)</th>
<th>MRL (3 labelled)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Effect of increasing the amount of labelled/unlabelled samples

→ As the size of the labelled set is reduced - performance gap increases
→ Locate the source even with few labelled samples, using unlabelled information
Why does Nearest-Neighbour fail?

Compare distances before and after mapping

Monotony
  - Before mapping - monotonic/ordered only in a limited region
  - After mapping - monotonic/ordered for almost the entire range
Why does Nearest-Neighbour fail?

Compare distances before and after mapping

Monotony
- Before mapping - monotonic/ordered only in a limited region
- After mapping - monotonic/ordered for almost the entire range

We conclude:
- RTFs lie on a nonlinear manifold - linear only for small patches
- NN ignores the manifold, MRL exploits the manifold structure
Outline

1. Data model and acoustic features
2. The Acoustic Manifold
3. Data-Driven Source Localization: Microphone Pair
4. Bayesian Perspective
5. Data-Driven Source Localization: Ad Hoc Array
Estimate the function $f$ which transforms an RTF to position using a Bayesian approach with a data-driven geometric model.

$$p(f|P_L, H_L) \propto p(P_L|f, H_L) \cdot p(f|H_L)$$

- **Posterior**: Correspondence between the function values and the labels
- **Likelihood**: A priori belief on the behavior of $f$ - controls smoothness
- **Prior**:

**Keywords**: Bayesian Perspective, Manifold-Based Bayesian Inference, [Laufer-Goldshtein et al., 2016a]
Estimate the function $f$ which transforms an RTF to position using a Bayesian approach with a data-driven geometric model.

$$p(f|P_L, H_L) \propto p(P_L|f, H_L) \cdot p(f|H_L)$$

**Posterior**

**Likelihood**

**Prior**

**Correspondence** between the function values and the labels

**A priori belief** on the behavior of $f$ - controls smoothness
Estimate the function $f$ which transforms an RTF to position using a Bayesian approach with a data-driven geometric model.

$$p(f|P_L, H_L) \propto p(P_L|f, H_L) \cdot p(f|H_L)$$

- **Posterior**
- **Likelihood**
- **Prior**

**Correspondence between the function values and the labels**

**A priori belief on the behavior of $f$ - controls smoothness**
Bayesian Perspective

Manifold-Based Bayesian Inference [Laufer-Goldshtein et al., 2016a]

Estimate the function $f$ which transforms an RTF to position using a Bayesian approach with a data-driven geometric model.

$$p(f|P_L, H_L) \propto p(P_L|f, H_L) \cdot p(f|H_L)$$

- **Posterior**
- **Likelihood**
- **Prior**

**Correspondence** between the function values and the labels

**A priori belief** on the behavior of $f$ - controls smoothness
Estimate the function $f$ which transforms an RTF to position using a Bayesian approach with a data-driven geometric model.
Bayesian Perspective

Statistical Model

The model:

1. An RTF is sampled from the manifold $\mathcal{M}$
2. The function $f$ follows a stochastic process
3. The function receives an RTF sample and returns the position
4. Measure a noisy position due to imperfect calibration
Statistical Model

The model:
1. An RTF is sampled from the manifold \( \mathcal{M} \)
2. The function \( f \) follows a stochastic process
3. The function receives an RTF sample and returns the position
4. Measure a noisy position due to imperfect calibration

\[ p(P_L|f, H_L) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n_L} (\bar{p}_i - f(h_i))^2 \right\} \]

→ What about the prior?
Standard Prior Probability

Standard Gaussian process:

- The function $f$ follows a Gaussian process:
  \[
  f(h) \sim \mathcal{GP}(\nu(h), k(h, h_i))
  \]

- $\nu$ is the mean function (choose $\nu \equiv 0$).
- $k$ is the covariance function.
- The r.v. $f_H = [f(h_1), \ldots, f(h_n)]$ has a joint Gaussian distribution:
  \[
  f_H \sim \mathcal{N}(0_n, \Sigma_{HH})
  \]

  where $\Sigma_{HH}$ is the covariance matrix with elements $k(h_i, h_j)$

- Common choice is a Gaussian kernel $k(h_i, h_j) = \exp\{-\|h_i - h_j\|^2/\varepsilon_k\}$
Bayesian Perspective

Prior Probability

Standard Prior Probability

Standard Gaussian process:

- The function $f$ follows a Gaussian process:

$$f(h) \sim \mathcal{GP}(\nu(h), k(h, h_i))$$

- $\nu$ is the mean function (choose $\nu \equiv 0$).
- $k$ is the covariance function.
- The r.v. $f_H = [f(h_1), \ldots, f(h_n)]$ has a joint Gaussian distribution:

$$f_H \sim \mathcal{N}(0_n, \Sigma_{HH})$$

where $\Sigma_{HH}$ is the covariance matrix with elements $k(h_i, h_j)$

- Common choice is a Gaussian kernel $k(h_i, h_j) = \exp\{-\|h_i - h_j\|^2/\varepsilon_k\}$

- The correlation for intermediate distances may be incorrectly assessed.
- Does not exploit the available set of unlabelled data $H_U$
Manifold-Based Prior Probability  [Sindhwani et al., 2007]

Discretization of the manifold

- The manifold is empirically represented by a graph $G$, with weights:
  $$W_{ij} = \begin{cases} 
  \exp \left\{ - \frac{\|h_i - h_j\|^2}{\varepsilon_w} \right\} & \text{if } h_j \in \mathcal{N}_i \text{ or } h_i \in \mathcal{N}_j \\
  0 & \text{otherwise}
  \end{cases}$$

- The graph Laplacian of $G$: $M = S - W$, where $S_{ii} = \sum_{j=1}^{n} W_{ij}$.
Bayesian Perspective  Prior Probability

Manifold-Based Prior Probability [Sindhwani et al., 2007]

Discretization of the manifold

- The manifold is empirically represented by a graph $G$, with weights:

$$W_{ij} = \begin{cases} \exp \left\{ - \frac{||h_i - h_j||^2}{\varepsilon_w} \right\} & \text{if } h_j \in \mathcal{N}_i \text{ or } h_i \in \mathcal{N}_j \\ 0 & \text{otherwise} \end{cases}$$

- The graph Laplacian of $G$: $M = S - W$, where $S_{ii} = \sum_{j=1}^{n} W_{ij}$.

Statistical formulation

- Geometry variables $G$ - represent the manifold structure

- The likelihood of $G$:

$$P(G|f_D) \propto \exp \left\{ - \frac{\gamma M}{2} \left( f_D^T M f_D \right) \right\}$$

- It can be shown:

$$f_D^T M f_D = \frac{1}{2} \sum_{i,j=1}^{n_D} W_{ij} (f(h_i) - f(h_j))^2$$
Manifold-Based Prior Probability \cite{Sindhwani et al., 2007}

**Manifold-Based GP Prior**

\[ p(f_H | G) \]

The covariance is formed by a manifold-based kernel \( \tilde{k} \)

**Likelihood of Geometry Variables**

\[ p(G | f_H) \]

correspondence between the function values and the manifold structure

Assume:

\[ p(G | f_H) \propto p(G | f_D) \]

\[ p(f_H | G) = \mathcal{N}(0_m, \tilde{\Sigma}_{HH}) \]

\[ P(G | f_D) \propto \exp \left\{ -\frac{\gamma M}{2} (f_D^T M f_D) \right\} \]

\[ p(f_H) = \mathcal{N}(0_m, \Sigma_{HH}) \]

**Standard GP Prior**

\[ p(f_H) \]

The covariance is formed by a standard kernel \( k \)

\[ \tilde{\Sigma}_{HH} \leftrightarrow \tilde{k}(h_i, h_j) \]
Bayesian Perspective

Localization

MAP/MMSE estimator:

- **Goal:** estimate the function value at some test sample $h_t \in M$.
- The training positions $\bar{p}_L = \text{vec}\{P_L\}$ and $f(h_t)$ are jointly Gaussian:

\[
\begin{bmatrix}
\bar{p}_L \\
\end{bmatrix}
| H_L, H_U \sim \mathcal{N}
\left(0_{nL+1},
\begin{bmatrix}
\tilde{\Sigma}_{LL} + \sigma^2 I_{nL} & \tilde{\Sigma}_{Lt} \\
\tilde{\Sigma}_{Tt} & \tilde{\Sigma}_{tt}
\end{bmatrix}
\right)
\]

- The posterior $p(f(h_t)|P_L, H_L, H_U)$ is a multivariate Gaussian with:

\[
\mu_{\text{cond}} = \tilde{\Sigma}_{Lt} \left(\tilde{\Sigma}_{LL} + \sigma^2 I_{nL}\right)^{-1} \bar{p}_L
\]

\[
\sigma^2_{\text{cond}} = \tilde{\Sigma}_{tt} - \tilde{\Sigma}_{Lt} \left(\tilde{\Sigma}_{LL} + \sigma^2 I_{nL}\right)^{-1} \tilde{\Sigma}_{Lt}
\]

The MAP/MMSE estimator of $f(h_t)$ is given by:

\[
\hat{f}(h_t) = \mu_{\text{cond}} = \tilde{\Sigma}_{Lt} \left(\tilde{\Sigma}_{LL} + \sigma^2 I_{nL}\right)^{-1} \bar{p}_L
\]
Analogy to Manifold Regularization for Localization

Formulate the estimation of $f$ as a regularized optimization in a reproducing kernel Hilbert space (RKHS) [Laufer-Goldshtein et al., 2016c]
Analogy to Manifold Regularization for Localization

Formulate the estimation of $f$ as a \textit{regularized optimization} in a reproducing kernel Hilbert space (RKHS) [Laufer-Goldshtein et al., 2016c]

$$f^* = \arg \min_{f \in \mathcal{H}_k} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(h_i))^2 + \gamma_k \|f\|_{\mathcal{H}_k}^2 + \gamma_M f_D^T M f_D$$

- **Search in RKHS defined by the kernel $\mathcal{H}_k$**
- **Cost Function**
- **$\mathcal{H}_k$ norm**
- **Manifold Regularization**
Formulate the estimation of $f$ as a regularized optimization in a reproducing kernel Hilbert space (RKHS) \cite{Laufer-Goldshtein et al., 2016c}

$$f^* = \arg\min_{f \in \mathcal{H}_k} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(h_i))^2 + \gamma_k \|f\|_{\mathcal{H}_k}^2 + \gamma_M f^T D M f_D$$

Search in RKHS defined by the kernel $k$

Cost Function

$\mathcal{H}_k$ norm

Manifold Regularization

$$f^* = \arg\min_{f \in \tilde{\mathcal{H}}_k} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(h_i))^2 + \gamma_k \|f\|_{\mathcal{H}_{\tilde{k}}}^2$$

Search in RKHS defined by the kernel $\tilde{k}$

Cost Function

$\mathcal{H}_{\tilde{k}}$ norm
Bayesian Perspective

Localization

Analogy to Manifold Regularization for Localization

Formulate the estimation of $f$ as a regularized optimization in a reproducing kernel Hilbert space (RKHS) \cite{LauferGoldshtein2016c}

$$f^* = \arg\min_{f \in \mathcal{H}_k} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(h_i))^2 + \gamma_k \|f\|_{\mathcal{H}_k}^2 + \gamma_M f^T D M f_D$$

Search in RKHS defined by the kernel $\tilde{k}$

Cost Function

$\mathcal{H}_k$ norm

Manifold Regularization

Manifold-Based Prior

Gaussian Process with Covariance $\tilde{k}$

Likelihood Function

S. Gannot (BIU)

Speaker Localization on Manifolds

LVA/ICA, 21.2.2017 46 / 62
Outline

1. Data model and acoustic features
2. The Acoustic Manifold
3. Data-Driven Source Localization: Microphone Pair
4. Bayesian Perspective
5. Data-Driven Source Localization: Ad Hoc Array
Can we extend the data-driven approach to localization with ad hoc array of multiple pairs of microphones?
Mapping RTFs to positions

Define $f_m : \mathcal{M}_m \mapsto \mathbb{R}$ - mapping the $m$th RTF to position

Each node:
- Represents a different viewpoint on the same acoustic event
- Induces relations between RTFs according to the associated manifold
Mapping RTFs to positions

Let \( \mathbf{h}_i = \left[ (\mathbf{h}_1^i)^T, (\mathbf{h}_2^i)^T, \ldots, (\mathbf{h}_M^i)^T \right]^T \in \bigcup_{m=1}^{M} \mathcal{M}_m \)

How to fuse the different views in a unified mapping \( f : \bigcup_{m=1}^{M} \mathcal{M}_m \mapsto \mathbb{R} \) ?
Inta-Manifold Relations

The mapping follows a Gaussian process $f^m(h^m) \sim \mathcal{GP}(0, \tilde{k}_m(h^m, h^m_i))$

Covariance function

Defined by a new manifold-based covariance function:

$$\text{cov} \left( f^m(h^m_r), f^m(h^m_l) \right) \equiv \tilde{k}_m(h^m_r, h^m_l) = \sum_{i=1}^{n_D} k_m(h^m_r, h^m_i) k_m(h^m_l, h^m_i)$$

$$= 2k_m(h^m_r, h^m_l) + \sum_{i=1}^{n_D} k_m(h^m_r, h^m_i) k_m(h^m_l, h^m_i)$$

where $k_m(h^m_r, h^m_i)$ is the covariance between the $m$-th node and the $i$-th node.
Inta-Manifold Relations

The mapping follows a Gaussian process $f^m(h^m) \sim \mathcal{GP}(0, \tilde{k}_m(h^m, h^m_i))$

Covariance function

Defined by a new manifold-based covariance function:

$$\text{cov} \left( f^m(h^m_r), f^m(h^m_l) \right) \equiv \tilde{k}_m(h^m_r, h^m_l) = \sum_{i=1}^{n_D} k_m(h^m_r, h^m_i) k_m(h^m_l, h^m_i)$$

$$= 2k_m(h^m_r, h^m_l) + \sum_{i=1}^{n_D} \sum_{i \neq l, r} k_m(h^m_r, h^m_i) k_m(h^m_l, h^m_i)$$
Inta-Manifold Relations

The mapping follows a Gaussian process $f^m(h^m) \sim \mathcal{GP}(0, \tilde{k}_m(h^m, h_i^m))$

Covariance function

Defined by a new manifold-based covariance function:

$$\text{cov} (f^m(h^m_r), f^m(h^m_l)) \equiv \tilde{k}_m(h^m_r, h^m_l) = \sum_{i=1}^{n_D} k_m(h^m_r, h^m_i) k_m(h^m_l, h^m_i)$$

$$= 2k_m(h^m_r, h^m_l) + \sum_{i=1}^{n_D} k_m(h^m_r, h^m_i) k_m(h^m_l, h^m_i)$$
Inta-Manifold Relations

The mapping follows a Gaussian process $f^m(h^m) \sim \mathcal{GP}(0, \tilde{k}_m(h^m, h^m_i))$

Covariance function

Defined by a new manifold-based covariance function:

$$\text{cov}(f^m(h^m_r), f^m(h^m_l)) \equiv \tilde{k}_m(h^m_r, h^m_l) = \sum_{i=1}^{n_D} k_m(h^m_r, h^m_i)k_m(h^m_l, h^m_i)$$

$$= 2k_m(h^m_r, h^m_l) + \sum_{i=1, i\neq l,r}^{n_D} k_m(h^m_r, h^m_i)k_m(h^m_l, h^m_i)$$
**Inta-Manifold Relations**

The mapping follows a Gaussian process $f^m(h^m) \sim \mathcal{GP}(0, \tilde{k}_m(h^m, h^m_i))$

**Covariance function**

Defined by a new manifold-based covariance function:

$$\text{cov} \left( f^m(h^m_r), f^m(h^m_l) \right) \equiv \tilde{k}_m(h^m_r, h^m_l) = \sum_{i=1}^{n_D} k_m(h^m_r, h^m_i) k_m(h^m_l, h^m_i)$$

$$= 2k_m(h^m_r, h^m_l) + \sum_{i=1}^{n_D} k_m(h^m_r, h^m_l) k_m(h^m_i, h^m_i)$$
Inta-Manifold Relations

The mapping follows a Gaussian process $f^m(h^m) \sim \mathcal{GP}(0, \tilde{k}_m(h^m, h_i^m))$

Covariance function

Defined by a new manifold-based covariance function:

$$\text{cov}(f^m(h_r^m), f^m(h_l^m)) \equiv \tilde{k}_m(h_r^m, h_i^m) = \sum_{i=1}^{n_D} k_m(h_r^m, h_i^m) k_m(h_i^m, h_l^m)$$

$$= 2k_m(h_r^m, h_l^m) + \sum_{i=1, i \neq l,r}^{n_D} k_m(h_r^m, h_i^m) k_m(h_i^m, h_l^m)$$
Inter-Manifold Relations

How to measure relations between RTFs from different nodes?

Multi-node covariance

The covariance between $f^q(h^q_r)$ and $f^w(h^w_r)$:

$$\text{cov} \left( f^q(h^q_r), f^w(h^w_r) \right) = \sum_{i=1}^{n_D} k_q(h^q_r, h^q_i) k_w(h^w_i, h^w_r)$$

S. Gannot (BIU)
**Multiple Manifold Gaussian Process (MMGP)**

**MMGP**
- Define $f(h)$ as the mean of the $M$ Gaussian processes:
  \[
  f(h) = \frac{1}{M} \left( f^1(h) + f^2(h) + \ldots + f^M(h) \right)
  \]
- If the process are jointly Gaussian
  \[
  \to f(h) \sim \mathcal{GP}(0, \tilde{k}(h, h_i)).
  \]

**The covariance function of $p$**

\[
\text{cov} (f(h_r), f(h_l)) = \tilde{k}(h_r, h_l) = \frac{1}{M^2} \text{cov} \left( \sum_{q=1}^{M} f^q(h^q_r), \sum_{w=1}^{M} f^w(h^w_l) \right)
\]
\[
= \frac{1}{M^2} \sum_{q,w=1}^{M} \text{cov}(f^q(h^q_r), f^w(h^w_l)) = \frac{1}{M^2} \sum_{i=1}^{n_D} \sum_{q,w=1}^{M} k_q(h^q_r, h^q_i) k_w(h^w_l, h^w_i)
\]
Localization

**MAP/MMSE estimator:**

- **Goal:** estimate the function value at some test sample
  \[ h_t = \left[ (h_1^T), (h_2^T), \ldots, (h_M^T) \right]^T \in \bigcup_{m=1}^M M_m. \]
- **The training positions** \( \bar{p}_L = \text{vec}\{P_L\} \) and \( f(h_t) \) are jointly Gaussian:
  \[
  \begin{bmatrix}
  \bar{p}_L \\
  f(h_t)
  \end{bmatrix} | H_L, H_U \sim \mathcal{N} \left( 0_{nL+1}, \begin{bmatrix}
  \tilde{\Sigma}_{LL} + \sigma^2 I_{nL} & \tilde{\Sigma}_{Lt} \\
  \tilde{\Sigma}_{Lt}^T & \tilde{\Sigma}_{tt}
  \end{bmatrix} \right)
  \]
- **The posterior** \( P(f(h_t)|P_L, H_L, H_U) \) is a multivariate Gaussian with:
  \[
  \mu_{\text{cond}} = \tilde{\Sigma}_{Lt}^T \left( \tilde{\Sigma}_{LL} + \sigma^2 I_{nL} \right)^{-1} \bar{p}_L
  \]
  \[
  \sigma^2_{\text{cond}} = \tilde{\Sigma}_{tt} - \tilde{\Sigma}_{Lt}^T \left( \tilde{\Sigma}_{LL} + \sigma^2 I_{nL} \right)^{-1} \tilde{\Sigma}_{Lt}.
  \]
  \[
  \hat{f}(h_t) = \mu_{\text{cond}} = \tilde{\Sigma}_{Lt}^T \left( \tilde{\Sigma}_{LL} + \sigma^2 I_{nL} \right)^{-1} \bar{p}_L
  \]
Recordings Setup

Setup:

- Real recordings carried out at Bar-Ilan acoustic lab
- A $6 \times 6 \times 2.4\text{m}$ room controllable reverberation time (set to 620ms)
- Region of interest: Source position is confined to a $2.8 \times 2.1\text{m}$ area
- 3 microphone pairs with inter-distance of 0.2m (position unknown)
Recordings Setup

Setup:

- Real recordings carried out at Bar-Ilan acoustic lab
- A $6 \times 6 \times 2.4\text{m}$ room controllable reverberation time (set to 620ms)
- Region of interest: Source position is confined to a $2.8 \times 2.1\text{m}$ area
- 3 microphone pairs with inter-distance of 0.2m (position unknown)
Experimental Results [Laufer-Goldshtein et al., 2016d]

Setup:
- Training: 20 labelled samples (0.7m resolution), 50 unlabelled samples
- Test: 25 random samples in the defined region
- Two noise types: air-conditioner noise and babble noise

Compare with:
- Concatenated independent measurements (Kernel-mult)
- Average of single-node estimates (Mean)
- Beamformer scanning (SRP-PHAT [DiBiase et al., 2001])
Challenges and Perspectives

Summary

- Manifold learning approach for source localization
- Data-driven manifold inference
- Location is the controlling variable of the RTF manifold
- It’s practical!
- Active research field [Deleforge et al., 2015][Yu et al., 2016][Xiao et al., 2015]

Challenges

- What happens if the source moves? ⇒ Source tracking special session
  1 [Laufer-Goldshtein et al., 2017]
- Can we apply the approach for multiple concurrent speakers?
- Beamforming is more complicated as it targets enhanced speech rather than its location. Can we extend the approach?
  - A first attempt using projections to the inferred manifold [Talmon and Gannot, 2013]
Manifold Regularization

Measuring smoothness over $\mathcal{M}$:

- The gradient $\nabla_\mathcal{M} f(h)$ represents variations around $h$
- A natural choice for intrinsic regularization:

$$\|f\|_\mathcal{M}^2 = \int_{\mathcal{M}} \|\nabla_\mathcal{M} f(h)\|^2 dp(h)$$

which is a global measure of smoothness for $f$

- Stokes’ theorem links gradient and Laplacian:

$$\int_{\mathcal{M}} \|\nabla_\mathcal{M} f(h)\|^2 dp(h) = \int_{\mathcal{M}} f(h) \Delta_\mathcal{M} f(h) dp(h) = \langle f(h), \Delta_\mathcal{M} f(h) \rangle$$

where $\Delta_\mathcal{M}$ is the Laplace-Beltrami operator.

How to reconstruct the Laplace-Beltrami operator on $\mathcal{M}$ given the training samples from the manifold?
Graph Laplacian:

- The manifold is empirically represented by a graph $G$, with weights:
  \[ W_{ij} = \begin{cases} \exp \left\{ -\frac{\|h_i - h_j\|^2}{\varepsilon_w} \right\} & \text{if } h_j \in \mathcal{N}_i \text{ or } h_i \in \mathcal{N}_j \\ 0 & \text{otherwise} \end{cases} \]

  where $\mathcal{N}_j$ is a set consisting of the $d$ nearest-neighbours of $h_j$.

- The graph Laplacian of $G$: $M = S - W$, where $S_{ii} = \sum_{j=1}^{n} W_{ij}$.

- Smoothness functional of $G$:
  \[ \langle f_D, Mf_D \rangle = f_D^T M f_D \]

  where $f_D = [f(h_1), \ldots, f(h_{n_D})]$.

- It can be shown:
  \[ f_D^T M f_D = \frac{1}{2} \sum_{i,j=1}^{n_D} W_{ij} |f(h_i) - f(h_j)|^2 \]
References

  Image method for efficiently simulating small-room acoustics.

  Manifold regularization: A geometric framework for learning from labeled and unlabeled examples.

  Diffusion maps.

  Acoustic noise analysis and speech enhancement techniques for mobile radio application.

  Acoustic space learning for sound-source separation and localization on binaural manifolds.
  International journal of neural systems, 25(1).

  Robust localization in reverberant rooms.
  In Microphone Arrays, pages 157–180. Springer.

  Signal enhancement using beamforming and nonstationarity with applications to speech.

  Generating sensor signals in isotropic noise fields.
## References (cont.)

Analysis and synthesis of room reverberation based on a statistical time-frequency model.
In *Audio Engineering Society Convention 103*. Audio Engineering Society.

The generalized correlation method for estimation of time delay.

Spatial source subtraction based on incomplete measurements of relative transfer function.

Study on manifolds of acoustic responses.

Manifold-based Bayesian inference for semi-supervised source localization.
In *IEEE International Conference on Audio and Acoustic Signal Processing (ICASSP)*, Shanghai, China.

A real-life experimental study on semi-supervised source localization based on manifold regularization.
In *International conference on the science of electrical engineering (ICSEE)*, Eilat, Israel.

Semi-supervised sound source localization based on manifold regularization.

Semi-supervised source localization on multiple-manifolds with distributed microphones.

---

S. Gannot (BIU)

Speaker Localization on Manifolds

LVA/ICA, 21.2.2017  60 / 62
Speaker tracking on multiple-manifolds with distributed microphones.

Performance analysis of the covariance subtraction method for relative transfer function estimation and comparison to the covariance whitening method.

Simulating the response of multiple microphones to a single acoustic source in a reverberant room.

Playing billiards in the concert hall: The mathematical foundations of geometrical room acoustics.

The “schroeder frequency” revisited.

Semi-supervised gaussian process classifiers.

Supervised source localization using diffusion kernels.
In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 245–248.
Relative transfer function identification on manifolds for supervised GSC beamformers.
In *21st European Signal Processing Conference (EUSIPCO)*, Marrakech, Morocco.

A learning-based approach to direction of arrival estimation in noisy and reverberant environments.
In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 2814–2818. IEEE.

Localization based stereo speech source separation using probabilistic time-frequency masking and deep neural networks.