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

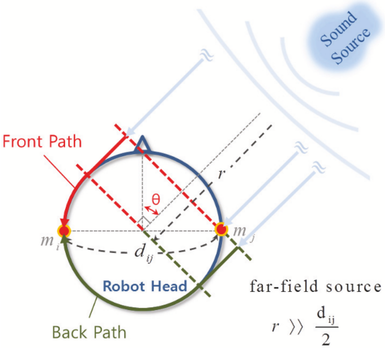
Speech enhancement tasks

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

Speaker Localization

Why localizing?

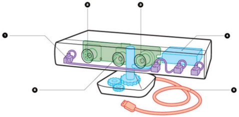
1. An essential component in many speech enhancement algorithms
2. Camera steering
3. Robot audition
4. Simultaneous localization and mapping (SLAM)



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

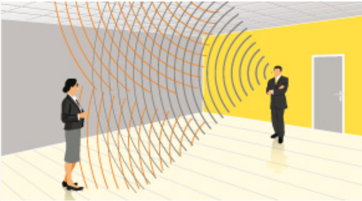
1. Cellular phones
2. Laptops and tablets
3. Hearing devices
4. Smart watches
5. Smart glasses
6. Smart homes & cars

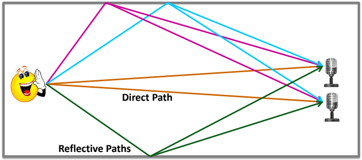


Room Acoustics Essentials

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]
- Statistical models for late reflections [Polack, 1993, Schroeder, 1996, Jot et al., 1997]
- 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

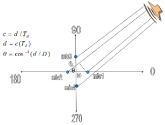
Data-driven approach

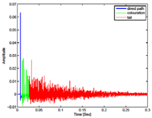
To alleviate these limitations and to infer a mathematical model that is accurate, simple to describe and simple to implement, we propose a data-driven approach

How to model the Acoustic Environment?

How to model the acoustic path?

- **TDOA-only** → Too simplified
- **Manifold Learning** → Interesting...
- **Entire Impulse Response** → Perfect but complicated

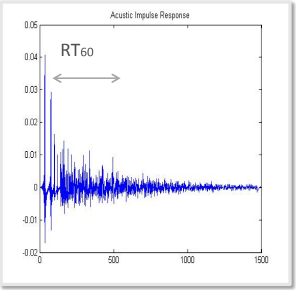




Main Claims

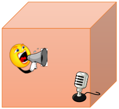
- The variability of the acoustic response in specific enclosures depends only on a small number of parameters
- The intrinsic degrees of freedom in acoustic responses are limited to a small number of variables

$\Rightarrow$ manifold learning approaches may improve localization ability



Controlling Parameters

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



Outline

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:

$$y_1(n) = a_1(n) * s(n) + u_1(n)$$
$$y_2(n) = a_2(n) * s(n) + u_2(n)$$

- $s(n)$ - 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:

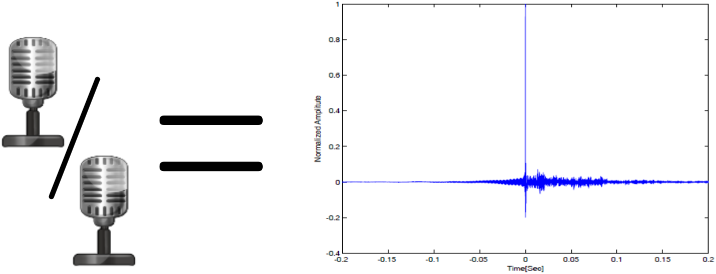
The measured signals in the two microphones:

$$y_1(n) = a_1(n) * s(n) + u_1(n)$$
$$y_2(n) = a_2(n) * s(n) + u_2(n)$$

- $s(n)$ - 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]



RTF:

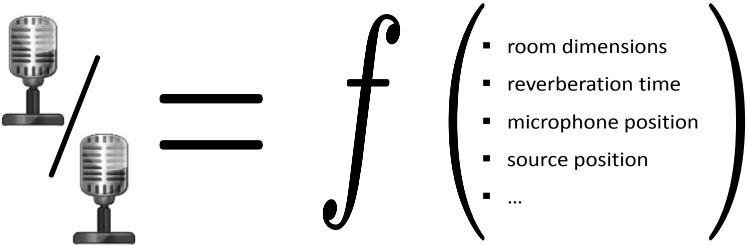
- Defined as the ratio between the transfer functions of the two mics:

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

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

$$a_2(n) = h_{12}(n) * a_1(n)$$

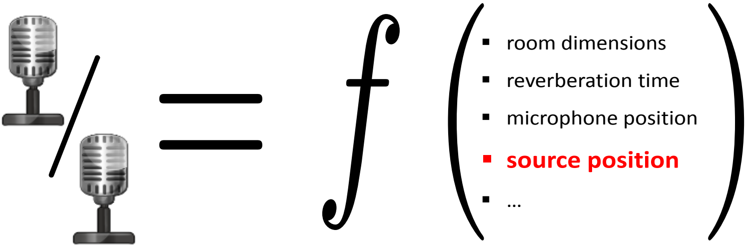
Relative Transfer Function (RTF) [Gannot et al., 2001]



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) [Gannot et al., 2001]



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

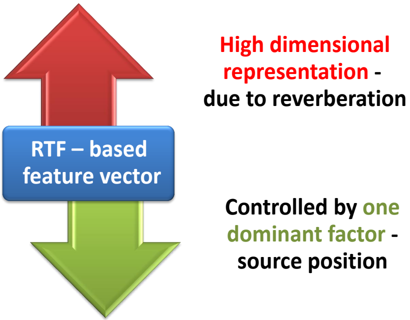
  (alternatively [Markovich-Golan and Gannot, 2015, Koldovsky et al., 2014]):

$$\hat{H}_{12}(k) = \frac{\hat{S}_{y_2 y_1}(k)}{\hat{S}_{y_1 y_1}(k)} \simeq \frac{A_2(k)}{A_1(k)}$$

- Define the feature vector:

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

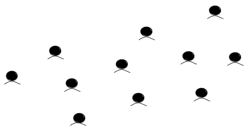
- $D \propto$ length of the relative impulse response (time domain)



Outline

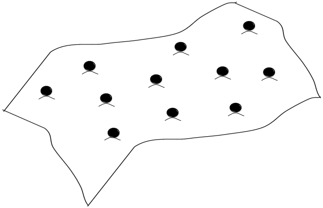
How to Measure the Affinity between RTFs? [Laufer-Goldshtein et al., 2015, Laufer-Goldshtein et al., 2016b]

- 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

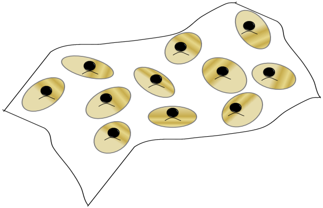


Acoustic manifold

- 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

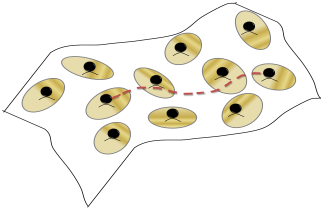


Acoustic manifold

- They lie on a low dimensional nonlinear manifold $\mathcal{M}$.
- Linearity is preserved in small neighbourhoods.

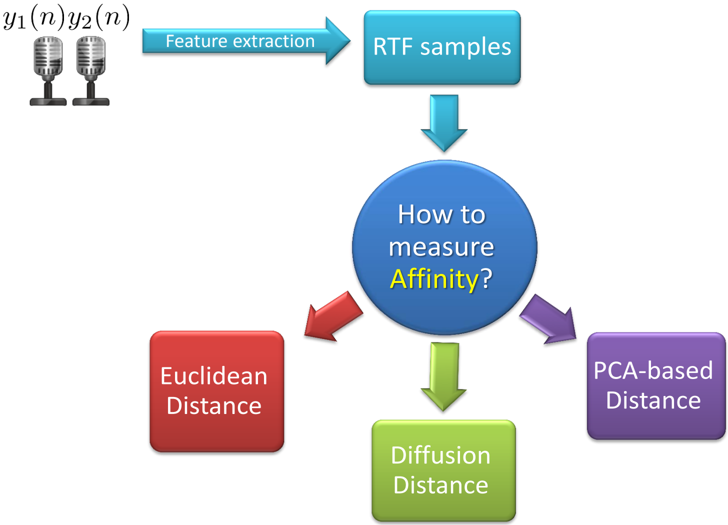
How to Measure the Affinity between RTFs? [Laufer-Goldshtein et al., 2015, Laufer-Goldshtein et al., 2016b]

- 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.
- Distances between RTFs should be measured along the manifold



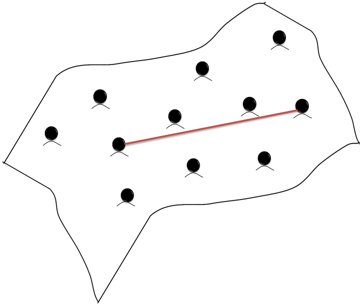
Each distance measure relies on a different hidden assumption about the underlying structure of the RTF samples

Euclidean Distance

The Euclidean distance between RTFs

$$D_{\text{Euc}}(\mathbf{h}_i, \mathbf{h}_j) = \|\mathbf{h}_i - \mathbf{h}_j\|$$

- 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

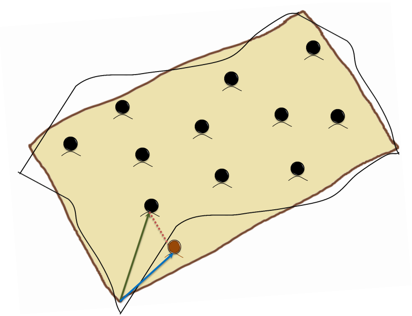
- The principal components - the d dominant eigenvectors $\{\mathbf{v}_i\}_{i=1}^{d}$ of the covariance matrix of the data
- The RTFs are linearly projected onto the principal components:

$$\boldsymbol{\nu}(\mathbf{h}_i) = [\mathbf{v}_1, \ldots \mathbf{v}_d]^T (\mathbf{h}_i - \boldsymbol{\mu}) .$$

PCA-based distance between RTFs

$$D_{\mathrm{PCA}}(\mathbf{h}_i, \mathbf{h}_j) = \|\boldsymbol{\nu}(\mathbf{h}_i) - \boldsymbol{\nu}(\mathbf{h}_j)\|.$$

- 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

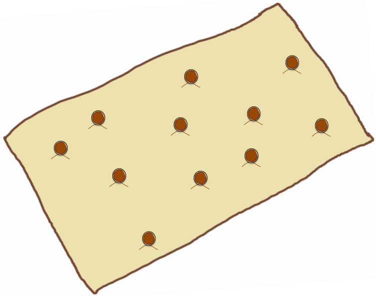
- The principal components - the d dominant eigenvectors $\{\mathbf{v}_i\}_{i=1}^{d}$ of the covariance matrix of the data
- The RTFs are linearly projected onto the principal components:

$$\boldsymbol{\nu}\left(\mathbf{h}_i\right) = \left[\mathbf{v}_1, \ldots \mathbf{v}_d\right]^T \left(\mathbf{h}_i - \boldsymbol{\mu}\right).$$

PCA-based distance between RTFs

$$D_{\mathrm{PCA}}(\mathbf{h}_i, \mathbf{h}_j) = \|\boldsymbol{\nu}(\mathbf{h}_i) - \boldsymbol{\nu}(\mathbf{h}_j)\|.$$

- 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

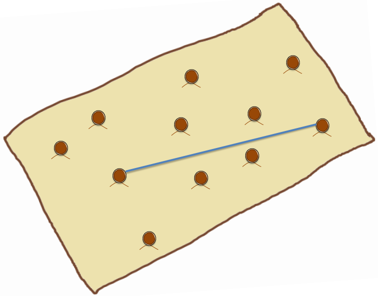
- The principal components - the d dominant eigenvectors $\{\mathbf{v}_i\}_{i=1}^{d}$ of the covariance matrix of the data
- The RTFs are linearly projected onto the principal components:

$$\boldsymbol{\nu}(\mathbf{h}_i) = [\mathbf{v}_1, \ldots \mathbf{v}_d]^T (\mathbf{h}_i - \boldsymbol{\mu}).$$

PCA-based distance between RTFs

$$D_{\mathrm{PCA}}(\mathbf{h}_i, \mathbf{h}_j) = \|\boldsymbol{\nu}(\mathbf{h}_i) - \boldsymbol{\nu}(\mathbf{h}_j)\|.$$

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

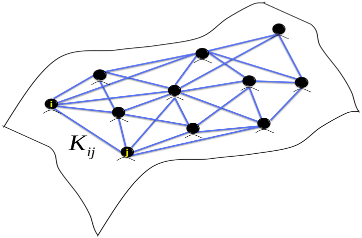


Diffusion Maps

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:

$$K_{ij} = k(\mathbf{h}_i, \mathbf{h}_j) = \exp\left\{-\frac{\|\mathbf{h}_i - \mathbf{h}_j\|^2}{\varepsilon}\right\}.$$

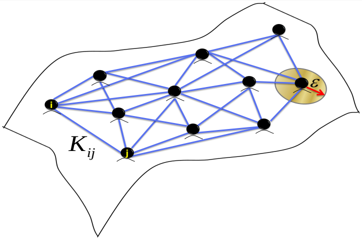


Diffusion Maps

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:

$$K_{ij} = k(\mathbf{h}_i, \mathbf{h}_j) = \exp\left\{-\frac{\|\mathbf{h}_i - \mathbf{h}_j\|^2}{\varepsilon}\right\}.$$



Diffusion Maps

Discretization of the manifold

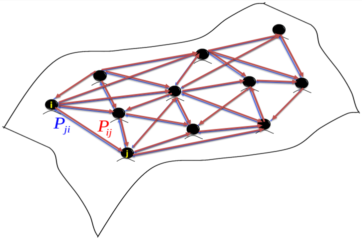
- The manifold can be empirically represented by a graph:
  - The RTF samples are the graph nodes.
  - The weights of the edges are defined using a kernel function:

$$K_{ij} = k(\mathbf{h}_i, \mathbf{h}_j) = \exp\left\{-\frac{\|\mathbf{h}_i - \mathbf{h}_j\|^2}{\varepsilon}\right\}.$$

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

$$p(\mathbf{h}_i, \mathbf{h}_j) = P_{ij} = K_{ij} / \sum_{r=1}^{N} K_{ir}.$$

which is a discretization of a diffusion process on the manifold

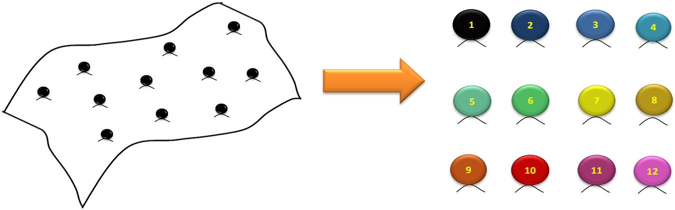


Diffusion Maps

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:

$$\mathbf{\Phi}_d : \mathbf{h}_i \mapsto \left[\lambda_1 \varphi_1^{(i)}, \ldots, \lambda_d \varphi_d^{(i)}\right]^T .$$



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

Diffusion Distance

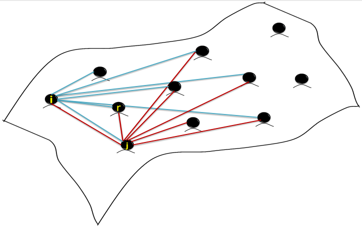
Diffusion distance between RTFs

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

$$D^2_{\mathrm{Diff}}(\mathbf{h}_i, \mathbf{h}_j) = \sum_{r=1}^{N} \left(p\left(\mathbf{h}_i, \mathbf{h}_r\right) - p\left(\mathbf{h}_j, \mathbf{h}_r\right) \right)^2 / \phi_0^{(r)}$$

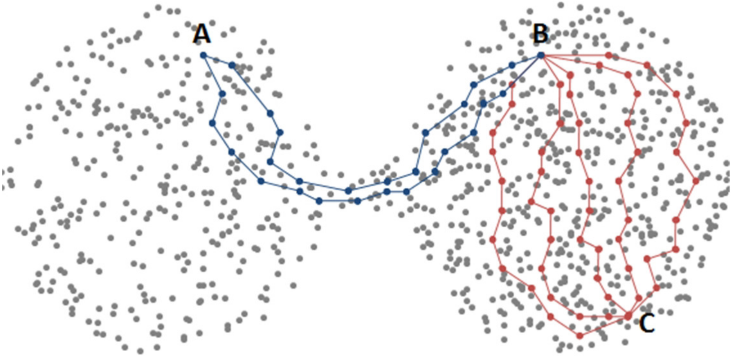
- Two points are close if they are highly connected in the graph.
- The diffusion distance can be well approximated by the Euclidian distance in the embedded domain:

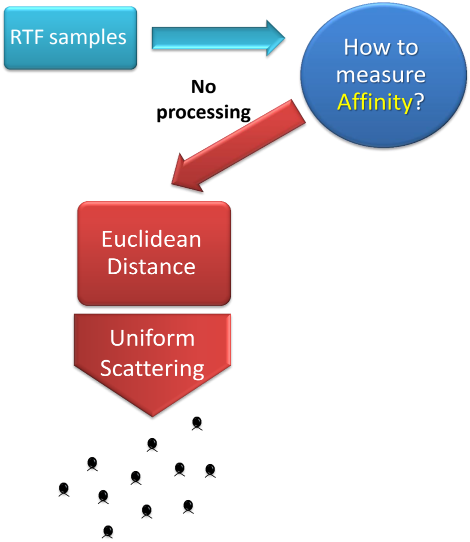
$$D_{\mathrm{Diff}}(\mathbf{h}_i, \mathbf{h}_j) \cong \|\mathbf{\Phi}_d(\mathbf{h}_i) - \mathbf{\Phi}_d(\mathbf{h}_j)\|.$$

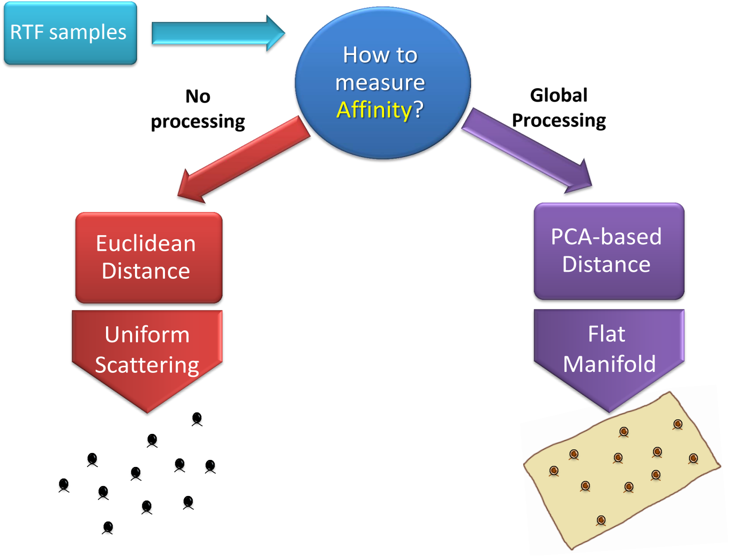


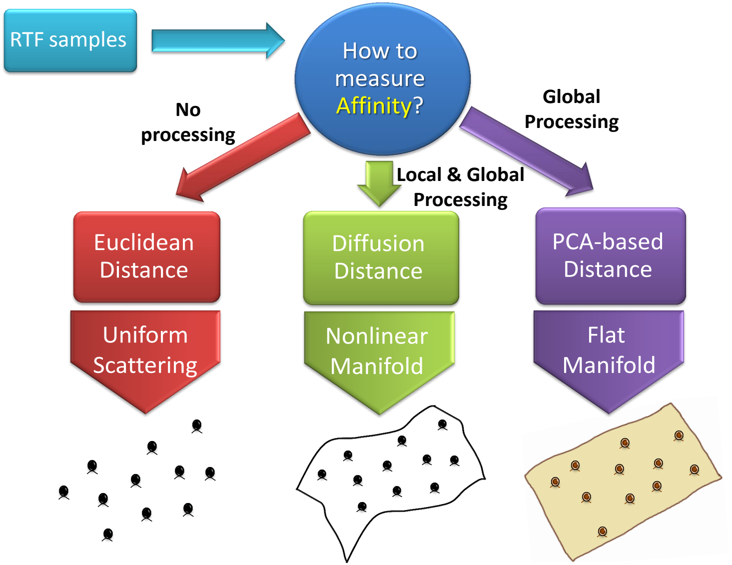
Diffusion Distance

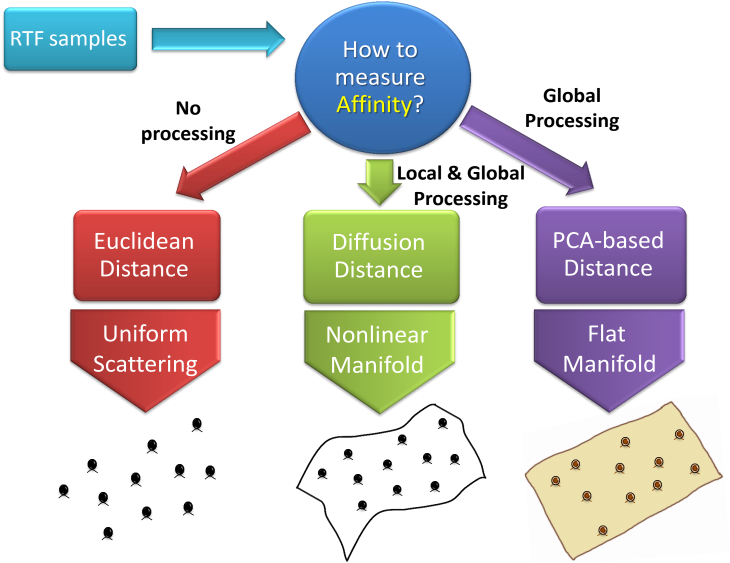
Illustration




RTF samples
How to measure Affinity?
No processing
Euclidean Distance
Uniform Scattering


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


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



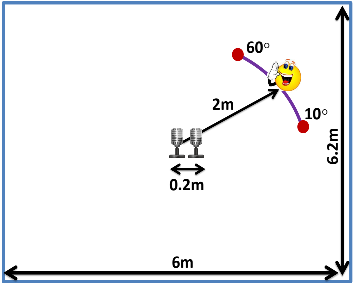
Which of the distance measures is proper?
What is the true underlying structure of the RTFs?

Simulation Results

Room setup

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

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

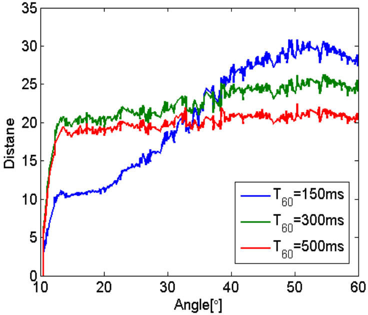


Test

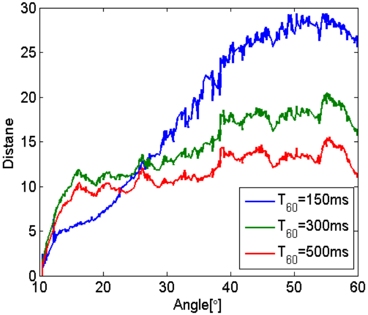
Measure the distance between each of the RTFs and the RTF corresponding to 10°:

- If monotonic with respect to the angle - proper distance
- If not monotonic with respect to the angle - improper distance

Euclidean Distance & PCA-based Distance [Laufer-Goldshtein et al., 2015]



(a) Euclidean Distance

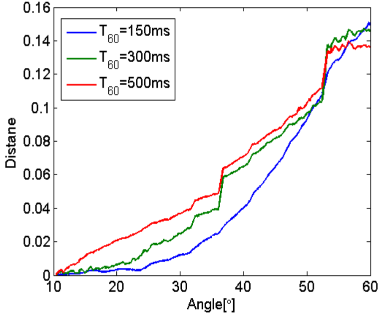


(b) PCA-based Distance

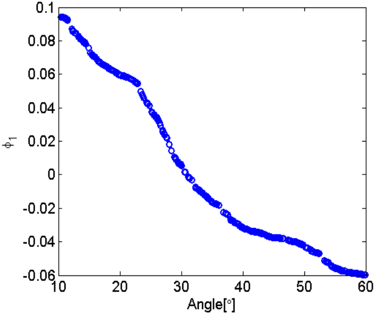
For both distance measures:

- 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

Diffusion Maps



(c) Diffusion Distance



(d) Diffusion Mapping

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

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

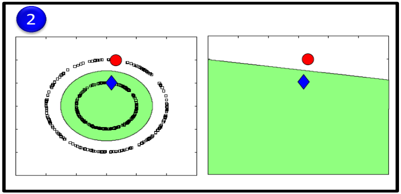
1

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

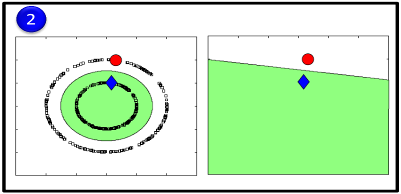


Semi-Supervised Learning

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

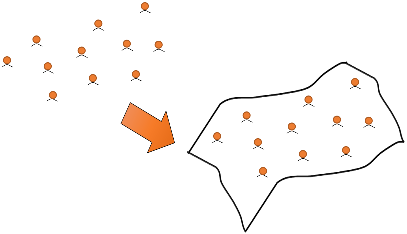
Why using unlabeled data?

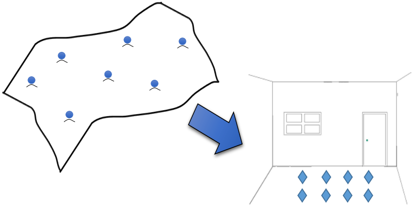
1. Localization - training should fit the specific environment of interest
   - Cannot generate a general database for all possible acoustic scenarios
   - Generating a large amount of labelled data is cumbersome/impractical
   - Unlabelled data is freely available - whenever someone is speaking
2. Unlabelled data can be utilize to recover the manifold structure
3. Semi-supervised learning is the natural setting for human learning



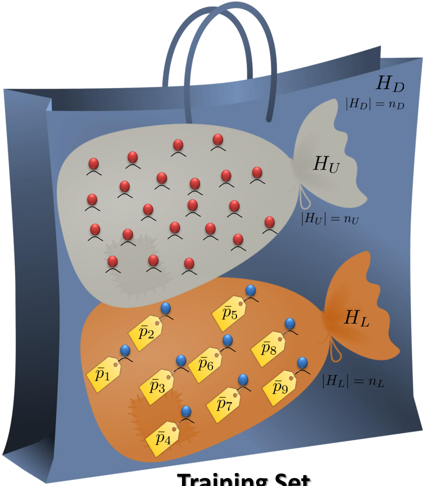
Semi-Supervised Learning

Unlabelled Samples	Labelled Samples
Recover the Manifold Structure	Anchor Points – Translate RTFs to Positions





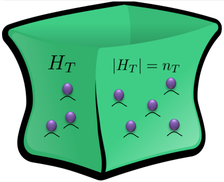
Semi-Supervised Learning



Training Set

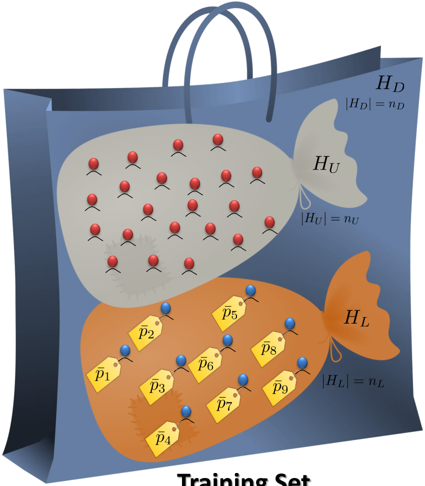
Data:

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



Test Set

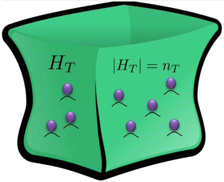
Semi-Supervised Learning



Training Set

Data:

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



Test Set

Goal: Recover the function f which transforms an RTF to position

Optimization and Manifold Regularization

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

$$f^* = \operatorname*{argmin}_{f \in \mathcal{H}_k} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \|f\|^2_{\mathcal{H}_k} + \gamma_M \|f\|^2_{\mathcal{M}}$$

Optimization and Manifold Regularization

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

$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \|f\|^2_{\mathcal{H}_k} + \gamma_M \|f\|^2_{\mathcal{M}}$$

Cost function

$\frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2$

correspondence between function values and labels

Optimization and Manifold Regularization

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

$$f^* = \operatorname*{argmin}_{f \in \mathcal{H}_k} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \|f\|^2_{\mathcal{H}_k} + \gamma_M \|f\|^2_{\mathcal{M}}$$

Cost function

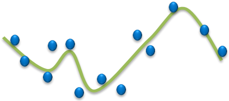
$\frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2$

Tikhonov Regularization

$\|f\|^2_{\mathcal{H}_k}$

correspondence between function values and labels

smoothness condition in the RKHS

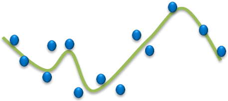
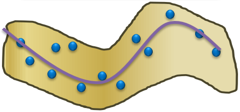


Optimization and Manifold Regularization

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

$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \|f\|_{\mathcal{H}_k}^2 + \gamma_M \|f\|_{\mathcal{M}}^2$$

Cost function	Tikhonov Regularization	Manifold Regularization
$\frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2$	$\|f\|_{\mathcal{H}_k}^2$	$\|f\|_{\mathcal{M}}^2$
correspondence between function values and labels	**smoothness condition in the RKHS**	**smoothness penalty with respect to the manifold**

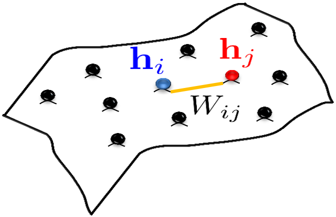
Manifold Regularization

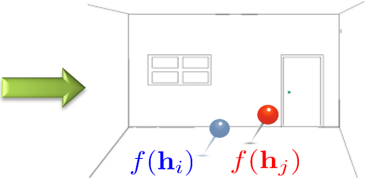
Discretization of the manifold

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

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

  where $\mathcal{N}_j$ is a set consisting of the d nearest-neighbours of $\mathbf{h}_j$.
- The graph Laplacian of G: $\mathbf{M} = \mathbf{S} - \mathbf{W}$, where $S_{ii} = \sum_{j=1}^{n_D} \mathbf{W}_{ij}$.
- Regularization: $\|f\|_{\mathcal{M}}^2 = \mathbf{f}_D^T \mathbf{M} \mathbf{f}_D = \frac{1}{2} \sum_{i,j=1}^{n_D} W_{ij} \left(f(\mathbf{h}_i) - f(\mathbf{h}_j)\right)^2$ with $\mathbf{f}_D^T = [f_1, f_2, \ldots, f_{n_D}]$ ▸ Proof





Optimization and Manifold Regularization

The optimization problem can be recast as:

$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \|f\|_{\mathcal{H}_k}^2 + \gamma_M \mathbf{f}_D^T \mathbf{M} \mathbf{f}_D$$

Optimization and Manifold Regularization

The optimization problem can be recast as:

$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \|f\|^2_{\mathcal{H}_k} + \gamma_M \mathbf{f}_D^T \mathbf{M} \mathbf{f}_D$$

The representer theorem:

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

$$f^*(\mathbf{h}) = \sum_{i=1}^{n_D} a_i k(\mathbf{h}_i, \mathbf{h})$$

where $k : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$ is the reproducing kernel of $\mathcal{H}_k$
with $K_{ij} = k(\mathbf{h}_i, \mathbf{h}_j) = \exp\left\{-\frac{\|\mathbf{h}_i - \mathbf{h}_j\|^2}{\varepsilon}\right\}$

Optimization and Manifold Regularization

The optimization problem can be recast as:

$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \|f\|^2_{\mathcal{H}_k} + \gamma_M \mathbf{f}_D^T \mathbf{M} \mathbf{f}_D$$

The representer theorem:

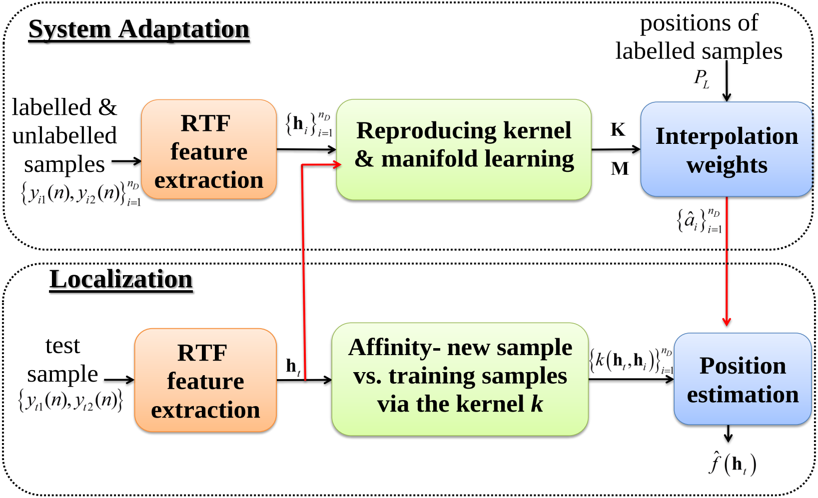
The minimizer over $\mathcal{H}_k$ of the regularized optimization is represented by:

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

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

Manifold Regularization for Localization (MRL)

[Laufer-Goldshtein et al., 2016c]

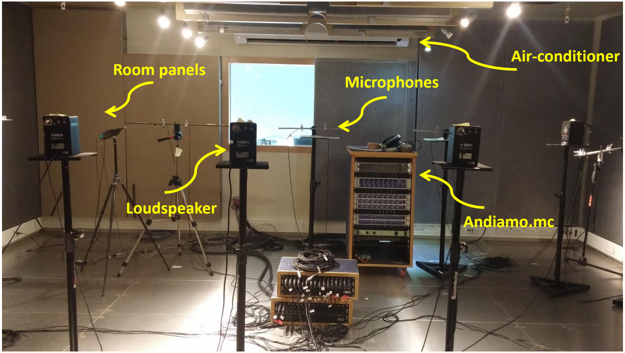


$K_{ij} = k(\mathbf{h}_i, \mathbf{h}_j)$

Recordings Setup

Setup:

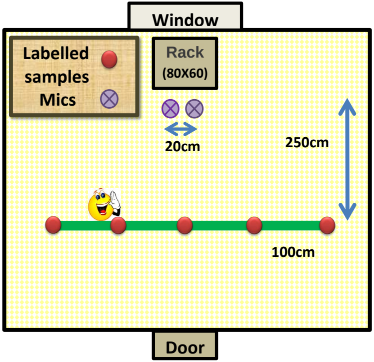
- Real recordings carried out at Bar-Ilan acoustic lab
- A $6 \times 6 \times 2.4$m room controllable reverberation time (set to 620ms)
- Region of interest: a 4m long line at 2.5m distance from the mics



Recordings Setup

Setup:

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



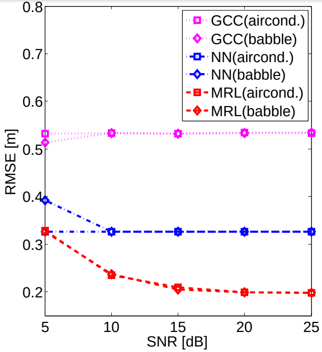
Experimental Results [Laufer-Goldshtein et al., 2016c]

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]



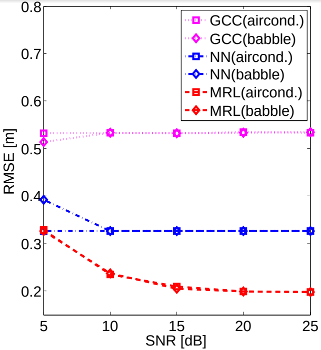
Experimental Results [Laufer-Goldshtein et al., 2016c]

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

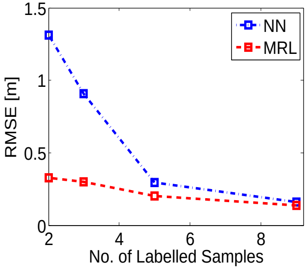
Compare with:

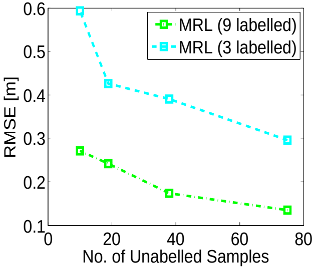
- Nearest-neighbour (NN)
- Generalized cross-correlation (GCC) method [Knapp and Carter, 1976]



The MRL algorithm outperforms the two other methods

Effect of Labelled & Unlabelled Samples





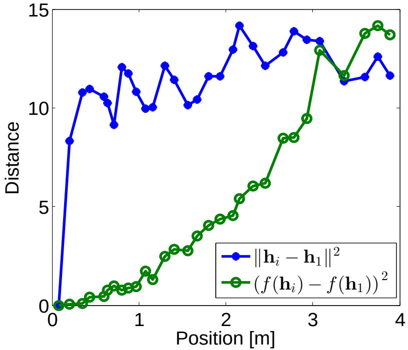
Effect of increasing the amount of labelled/unlabelled samples

→ 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

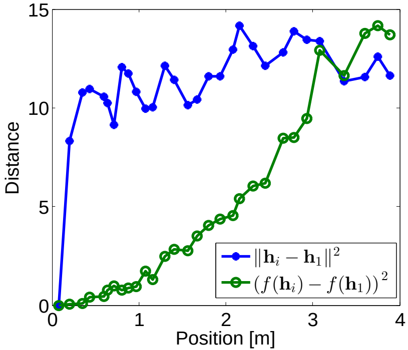


Monotony

- 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

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

Update Belief

$$\underbrace{p(f|P_L, H_L)}_{\text{Posterior}} \propto \underbrace{p(P_L|f, H_L)}_{\text{Likelihood}} \cdot \underbrace{p(f|H_L)}_{\text{Prior}}$$

Likelihood: Correspondence between the function values and the labels

Prior: A priori belief on the behavior of f - controls smoothness

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

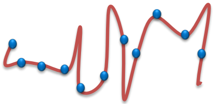
Update Belief

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

Posterior | **Likelihood** | **Prior**

Likelihood: Correspondence between the function values and the labels

Prior: A priori belief on the behavior of f - controls smoothness



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

Update Belief

$$\underbrace{p(f|P_L, H_L)}_{\text{Posterior}} \propto \underbrace{p(P_L|f, H_L)}_{\text{Likelihood}} \cdot \underbrace{p(f|H_L)}_{\text{Prior}}$$

Likelihood: Correspondence between the function values and the labels

Prior: A priori belief on the behavior of f - controls smoothness

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

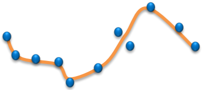
Estimate the function f which transforms an RTF to position using a Bayesian approach with a data-driven geometric model

Update Belief

$$\underbrace{p(f|P_L, H_L)}_{\text{Posterior}} \propto \underbrace{p(P_L|f, H_L)}_{\text{Likelihood}} \cdot \underbrace{p(f|H_L)}_{\text{Prior}}$$

Likelihood: Correspondence between the function values and the labels

Prior: A priori belief on the behavior of f - controls smoothness



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

Update Belief

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

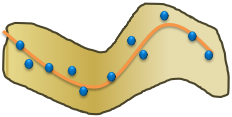
Posterior

Likelihood

Manifold-Based Prior

Correspondence between the function values and the labels

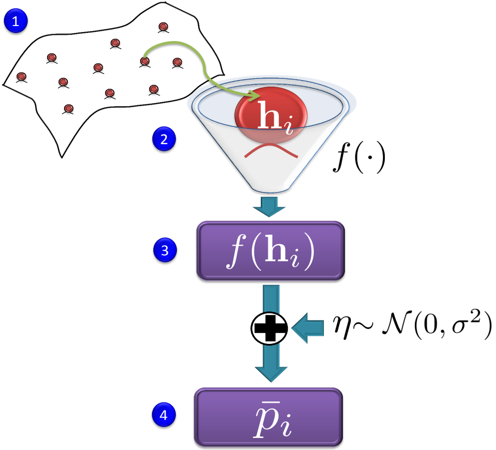
a priori belief on the properties of f - smoothness with respect to the manifold



Statistical Model

The model:

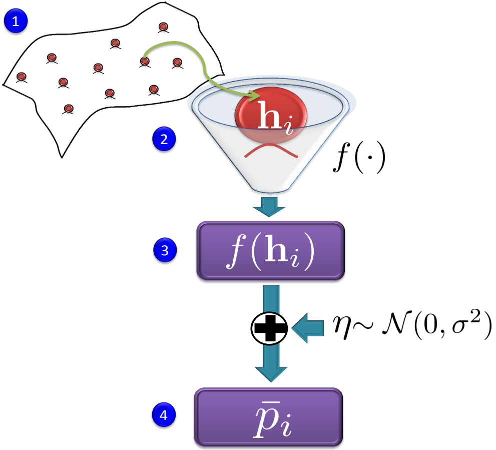
1. An RTF is sampled from the manifold $\mathcal{M}$
2. The function f follows a stochastic process
3. The function receives an RTF sample and returns the position
4. Measure a noisy position due to imperfect calibration



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



→ Likelihood function: $p(P_L|f, H_L) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n_L}(\bar{p}_i - f(\mathbf{h}_i))^2\right\}$

→ What about the prior?

Standard Prior Probability

Standard Gaussian process:

- The function f follows a Gaussian process:

$$f(\mathbf{h}) \sim \mathcal{GP}\left(\nu(\mathbf{h}), k(\mathbf{h}, \mathbf{h}_i)\right)$$

- ν is the mean function (choose $\nu \equiv 0$).
- k is the covariance function.
- The r.v. $\mathbf{f}_H = [f(\mathbf{h}_1), \ldots, f(\mathbf{h}_n)]$ has a joint Gaussian distribution:

$$\mathbf{f}_H \sim \mathcal{N}(\mathbf{0}_n, \mathbf{\Sigma}_{HH})$$

  where $\mathbf{\Sigma}_{HH}$ is the covariance matrix with elements $k(\mathbf{h}_i, \mathbf{h}_j)$
- Common choice is a Gaussian kernel $k(\mathbf{h}_i, \mathbf{h}_j) = \exp\{-\|\mathbf{h}_i - \mathbf{h}_j\|^2/\varepsilon_k\}$

Standard Prior Probability

Standard Gaussian process:

- The function f follows a Gaussian process:

$$f(\mathbf{h}) \sim \mathcal{GP}\left(\nu(\mathbf{h}), k(\mathbf{h}, \mathbf{h}_i)\right)$$

- ν is the mean function (choose $\nu \equiv 0$).
- k is the covariance function.
- The r.v. $\mathbf{f}_H = [f(\mathbf{h}_1), \ldots, f(\mathbf{h}_n)]$ has a joint Gaussian distribution:

$$\mathbf{f}_H \sim \mathcal{N}(\mathbf{0}_n, \mathbf{\Sigma}_{HH})$$

where $\mathbf{\Sigma}_{HH}$ is the covariance matrix with elements $k(\mathbf{h}_i, \mathbf{h}_j)$
- Common choice is a Gaussian kernel $k(\mathbf{h}_i, \mathbf{h}_j) = \exp\{-\|\mathbf{h}_i - \mathbf{h}_j\|^2 / \varepsilon_k\}$

✗ 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:

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

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

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

Discretization of the manifold

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

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

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

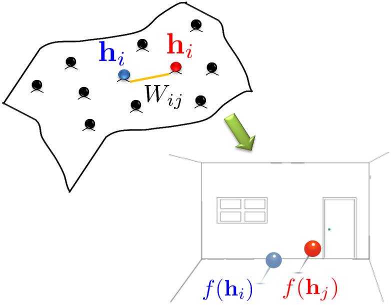
Statistical formulation

- Geometry variables $\mathcal{G}$ - represent the manifold structure
- The likelihood of $\mathcal{G}$:

$$P(\mathcal{G}|\mathbf{f}_D) \propto \exp\left\{-\frac{\gamma_M}{2}\left(\mathbf{f}_D^T \mathbf{M} \mathbf{f}_D\right)\right\}$$

- It can be shown:

$$\mathbf{f}_D^T \mathbf{M} \mathbf{f}_D = \frac{1}{2} \sum_{i,j=1}^{n_D} W_{ij} \left(f(\mathbf{h}_i) - f(\mathbf{h}_j)\right)^2$$



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

Manifold-Based GP Prior | **Likelihood of Geometry Variables** | **Standard GP Prior**

$$p(\mathbf{f}_H|\mathcal{G}) \propto p(\mathcal{G}|\mathbf{f}_H) \cdot p(\mathbf{f}_H)$$

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

correspondence between the function values and the manifold structure

The covariance is formed by a standard kernel k

Assume: $p(\mathcal{G}|\mathbf{f}_H) \propto p(\mathcal{G}|\mathbf{f}_D)$

$p(\mathbf{f}_H|\mathcal{G}) = \mathcal{N}(\mathbf{0}_m, \tilde{\mathbf{\Sigma}}_{HH})$

$P(\mathcal{G}|\mathbf{f}_D) \propto \exp\left\{ -\frac{\gamma_M}{2}\left(\mathbf{f}_D^T \mathbf{M} \mathbf{f}_D\right)\right\}$

$p(\mathbf{f}_H) = \mathcal{N}(\mathbf{0}_m, \mathbf{\Sigma}_{HH})$

$$\tilde{\Sigma}_{\mathrm{HH}} \Leftrightarrow \tilde{k}(\mathbf{h}_i, \mathbf{h}_j)$$

Localization

MAP/MMSE estimator:

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

$$\begin{bmatrix} \bar{\mathbf{p}}_L \\ f(\mathbf{h}_t) \end{bmatrix} \Big| H_L, H_U \sim \mathcal{N}\left(\mathbf{0}_{n_L+1}, \begin{bmatrix} \tilde{\mathbf{\Sigma}}_{LL} + \sigma^2 \mathbf{I}_{n_L} & \tilde{\mathbf{\Sigma}}_{Lt} \\ \tilde{\mathbf{\Sigma}}_{Lt}^T & \tilde{\mathbf{\Sigma}}_{tt} \end{bmatrix}\right)$$

- The posterior $p(f(\mathbf{h}_t)|P_L, H_L, H_U)$ is a multivariate Gaussian with:

$$\mu_{\text{cond}} = \tilde{\mathbf{\Sigma}}_{Lt}^T \left(\tilde{\mathbf{\Sigma}}_{LL} + \sigma^2 \mathbf{I}_{n_L}\right)^{-1} \bar{\mathbf{p}}_L$$

$$\sigma^2_{\text{cond}} = \tilde{\Sigma}_{tt} - \tilde{\mathbf{\Sigma}}_{Lt}^T \left(\tilde{\mathbf{\Sigma}}_{LL} + \sigma^2 \mathbf{I}_{n_L}\right)^{-1} \tilde{\mathbf{\Sigma}}_{Lt}.$$

The MAP/MMSE estimator of $f(\mathbf{h}_t)$ is given by:

$$\hat{f}(\mathbf{h}_t) = \mu_{\text{cond}} = \tilde{\mathbf{\Sigma}}_{Lt}^T \left(\tilde{\mathbf{\Sigma}}_{LL} + \sigma^2 \mathbf{I}_{n_L}\right)^{-1} \bar{\mathbf{p}}_L$$

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 regularized optimization in a reproducing kernel Hilbert space (RKHS) [Laufer-Goldshtein et al., 2016c]

$$f^* = \operatorname{argmin}_{f \in \mathcal{H}_k} \underbrace{\frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2}_{\text{Cost Function}} + \underbrace{\gamma_k \|f\|^2_{\mathcal{H}_k}}_{\mathcal{H}_k \text{ norm}} + \underbrace{\gamma_M \mathbf{f}_D^T \mathbf{M} \mathbf{f}_D}_{\text{Manifold Regularization}}$$

Search in RKHS defined by the kernel k

Cost Function

$\mathcal{H}_k$ norm

Manifold Regularization

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]

$$f^* = \operatorname{argmin}_{f \in \mathcal{H}_k} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \|f\|^2_{\mathcal{H}_k} + \gamma_M \mathbf{f}_D^T \mathbf{M} \mathbf{f}_D$$

Search in RKHS defined by the kernel k

Cost Function

$\mathcal{H}_k$ norm

Manifold Regularization

$$f^* = \operatorname{argmin}_{f \in \mathcal{H}_{\tilde{k}}} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \|f\|^2_{\mathcal{H}_{\tilde{k}}}$$

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

Cost Function

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

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]

$$f^* = \text{argmin}_{f \in \mathcal{H}_k} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \|f\|^2_{\mathcal{H}_k} + \gamma_M \mathbf{f}_D^T \mathbf{M} \mathbf{f}_D$$

- Search in RKHS defined by the kernel k
- Cost Function
- $\mathcal{H}_k$ norm
- Manifold Regularization

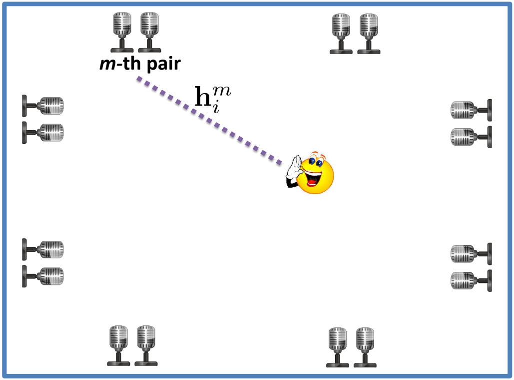
$$f^* = \text{argmin}_{f \in \mathcal{H}_{\tilde{k}}} \frac{1}{n_L} \sum_{i=1}^{n_L} (\bar{p}_i - f(\mathbf{h}_i))^2 + \gamma_k \|f\|^2_{\mathcal{H}_{\tilde{k}}}$$

- Search in RKHS defined by the kernel $\tilde{k}$ → Gaussian Process with Covariance $\tilde{k}$
- Cost Function → Likelihood Function
- $\mathcal{H}_{\tilde{k}}$ norm → Manifold-Based Prior

Outline

Source Localization with Ad Hoc Microphone Array

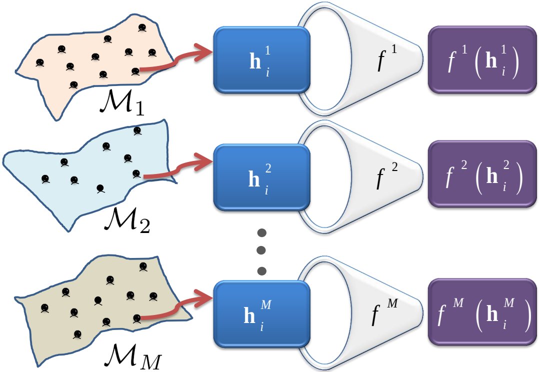
[Laufer-Goldshtein et al., 2016d]



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 mth RTF to position

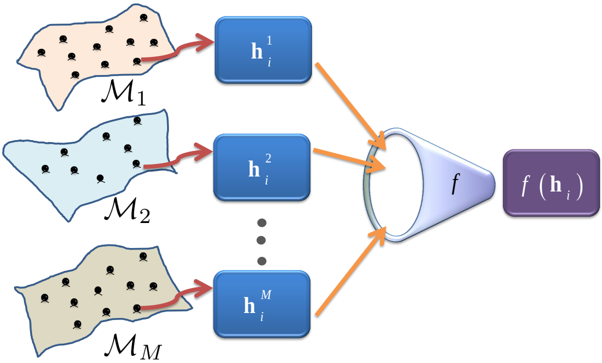


Each node:

- Represents a different view point on the same acoustic event
- Induces relations between RTFs according to the associated manifold

Mapping RTFs to positions

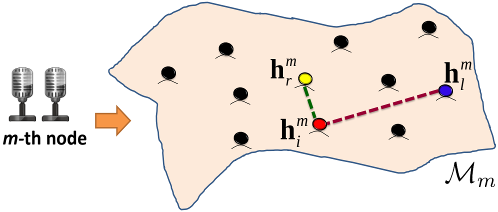
Let $\mathbf{h}_i = \left[(\mathbf{h}_i^1)^T, (\mathbf{h}_i^2)^T, \ldots, (\mathbf{h}_i^M)^T\right]^T \in \cup_{m=1}^{M} \mathcal{M}_m$



How to fuse the different views in a unified mapping $f : \cup_{m=1}^{M} \mathcal{M}_m \mapsto \mathbb{R}$?

Inta-Manifold Relations

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



Covariance function

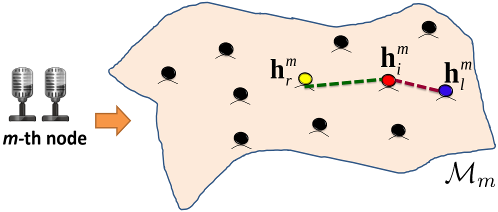
Defined by a new manifold-based covariance function:

$$\mathrm{cov}\left(f^m(\mathbf{h}_r^m), f^m(\mathbf{h}_l^m)\right) \equiv \tilde{k}_m(\mathbf{h}_r^m, \mathbf{h}_l^m) = \sum_{i=1}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

$$= 2k_m(\mathbf{h}_r^m, \mathbf{h}_l^m) + \sum_{\substack{i=1 \\ i \neq l,r}}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

Inta-Manifold Relations

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



Covariance function

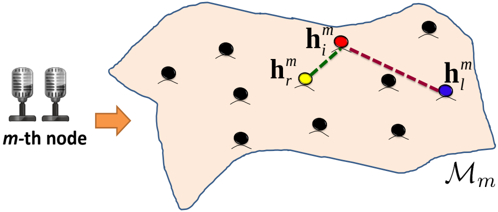
Defined by a new manifold-based covariance function:

$$\mathrm{cov}\left(f^m(\mathbf{h}_r^m), f^m(\mathbf{h}_l^m)\right) \equiv \tilde{k}_m(\mathbf{h}_r^m, \mathbf{h}_l^m) = \sum_{i=1}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

$$= 2k_m(\mathbf{h}_r^m, \mathbf{h}_l^m) + \sum_{\substack{i=1 \\ i \neq l,r}}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

Inta-Manifold Relations

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



Covariance function

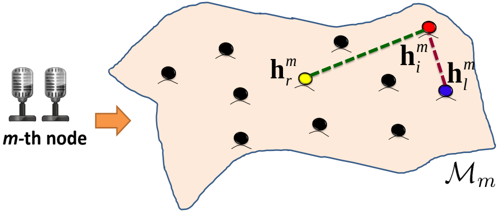
Defined by a new manifold-based covariance function:

$$\operatorname{cov}\left(f^m(\mathbf{h}_r^m), f^m(\mathbf{h}_l^m)\right) \equiv \tilde{k}_m(\mathbf{h}_r^m, \mathbf{h}_l^m) = \sum_{i=1}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

$$= 2k_m(\mathbf{h}_r^m, \mathbf{h}_l^m) + \sum_{\substack{i=1 \\ i \neq l,r}}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

Inta-Manifold Relations

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



Covariance function

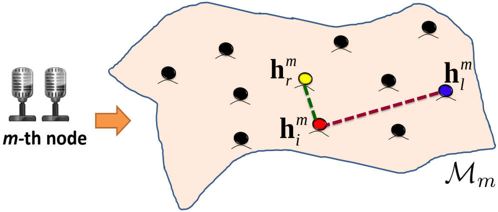
Defined by a new manifold-based covariance function:

$$\mathrm{cov}\left(f^m(\mathbf{h}_r^m), f^m(\mathbf{h}_l^m)\right) \equiv \tilde{k}_m(\mathbf{h}_r^m, \mathbf{h}_l^m) = \sum_{i=1}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

$$= 2k_m(\mathbf{h}_r^m, \mathbf{h}_l^m) + \sum_{\substack{i=1 \\ i \neq l,r}}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

Inta-Manifold Relations

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



Covariance function

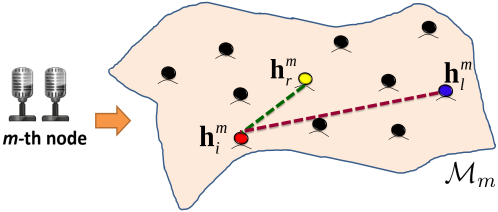
Defined by a new manifold-based covariance function:

$$\operatorname{cov}(f^m(\mathbf{h}_r^m), f^m(\mathbf{h}_l^m)) \equiv \tilde{k}_m(\mathbf{h}_r^m, \mathbf{h}_l^m) = \sum_{i=1}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

$$= 2k_m(\mathbf{h}_r^m, \mathbf{h}_l^m) + \sum_{\substack{i=1 \\ i \neq l,r}}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

Inta-Manifold Relations

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



Covariance function

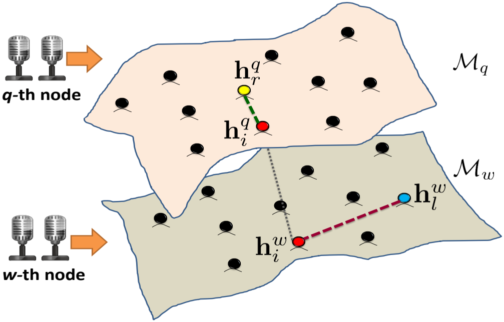
Defined by a new manifold-based covariance function:

$$\mathrm{cov}\left(f^m(\mathbf{h}_r^m), f^m(\mathbf{h}_l^m)\right) \equiv \tilde{k}_m(\mathbf{h}_r^m, \mathbf{h}_l^m) = \sum_{i=1}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

$$= 2k_m(\mathbf{h}_r^m, \mathbf{h}_l^m) + \sum_{\substack{i=1 \\ i \neq l,r}}^{n_D} k_m(\mathbf{h}_r^m, \mathbf{h}_i^m) k_m(\mathbf{h}_l^m, \mathbf{h}_i^m)$$

Inter-Manifold Relations

How to measure relations between RTFs from different nodes?



Multi-node covariance

The covariance between $f^q(\mathbf{h}_r^q)$ and $f^w(\mathbf{h}_r^w)$:

$$\mathrm{cov}\left(f^q(\mathbf{h}_r^q), f^w(\mathbf{h}_r^w)\right) = \sum_{i=1}^{n_D} k_q(\mathbf{h}_r^q, \mathbf{h}_i^q) k_w(\mathbf{h}_l^w, \mathbf{h}_i^w)$$

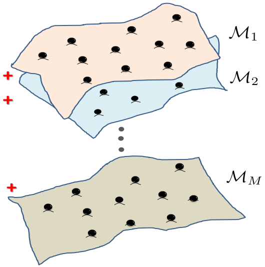
Multiple Manifold Gaussian Process (MMGP)

MMGP

- Define $f(\mathbf{h})$ as the mean of the M Gaussian processes:

$$f(\mathbf{h}) = \frac{1}{M}\left(f^1(\mathbf{h}) + f^2(\mathbf{h}) + \ldots + f^M(\mathbf{h})\right)$$

- If the process are jointly Gaussian $\rightarrow f(\mathbf{h}) \sim \mathcal{GP}(0, \tilde{k}(\mathbf{h}, \mathbf{h}_i))$.



The covariance function of p

$$\mathrm{cov}\left(f(\mathbf{h}_r), f(\mathbf{h}_l)\right) = \tilde{k}(\mathbf{h}_r, \mathbf{h}_l) = \frac{1}{M^2}\mathrm{cov}\left(\sum_{q=1}^{M} f^q(\mathbf{h}_r^q), \sum_{w=1}^{M} f^w(\mathbf{h}_l^w)\right)$$

$$= \frac{1}{M^2}\sum_{q,w=1}^{M}\mathrm{cov}(f^q(\mathbf{h}_r^q), f^w(\mathbf{h}_r^w)) = \frac{1}{M^2}\sum_{i=1}^{n_D}\sum_{q,w=1}^{M} k_q(\mathbf{h}_r^q, \mathbf{h}_i^q)k_w(\mathbf{h}_l^w, \mathbf{h}_i^w)$$

Localization

MAP/MMSE estimator:

- Goal: estimate the function value at some test sample $\mathbf{h}_t = \left[(\mathbf{h}_t^1)^T, (\mathbf{h}_t^2)T, \ldots, (\mathbf{h}_t^M)^T\right]^T \in \cup_{m=1}^M \mathcal{M}_m$.
- The training positions $\bar{\mathbf{p}}_L = \mathrm{vec}\{P_L\}$ and $f(\mathbf{h}_t)$ are jointly Gaussian:

$$\begin{bmatrix} \bar{\mathbf{p}}_L \\ f(\mathbf{h}_t) \end{bmatrix} \Big| H_L, H_U \sim \mathcal{N}\left(\mathbf{0}_{n_L+1}, \begin{bmatrix} \tilde{\mathbf{\Sigma}}_{LL} + \sigma^2 \mathbf{I}_{n_L} & \tilde{\mathbf{\Sigma}}_{Lt} \\ \tilde{\mathbf{\Sigma}}_{Lt}^T & \tilde{\mathbf{\Sigma}}_{tt} \end{bmatrix}\right)$$

- The posterior $P\left(f(\mathbf{h}_t) | P_L, H_L, H_U\right)$ is a multivariate Gaussian with:

$$\mu_{\mathrm{cond}} = \tilde{\mathbf{\Sigma}}_{Lt}^T \left(\tilde{\mathbf{\Sigma}}_{LL} + \sigma^2 \mathbf{I}_{n_L}\right)^{-1} \bar{\mathbf{p}}_L$$

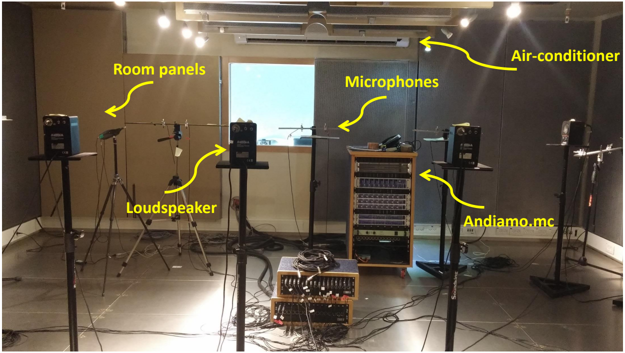
$$\sigma^2_{\mathrm{cond}} = \tilde{\Sigma}_{tt} - \tilde{\mathbf{\Sigma}}_{Lt}^T \left(\tilde{\mathbf{\Sigma}}_{LL} + \sigma^2 \mathbf{I}_{n_L}\right)^{-1} \tilde{\mathbf{\Sigma}}_{Lt}.$$

$$\hat{f}(\mathbf{h}_t) = \mu_{\mathrm{cond}} = \tilde{\mathbf{\Sigma}}_{Lt}^T \left(\tilde{\mathbf{\Sigma}}_{LL} + \sigma^2 \mathbf{I}_{n_L}\right)^{-1} \bar{\mathbf{p}}_L$$

Recordings Setup

Setup:

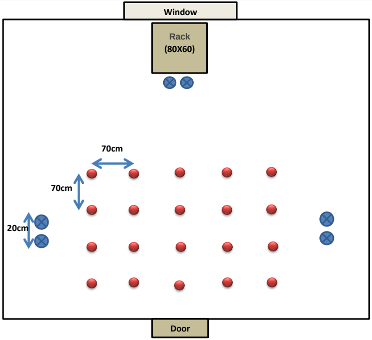
- Real recordings carried out at Bar-Ilan acoustic lab
- A $6 \times 6 \times 2.4$m room controllable reverberation time (set to 620ms)
- Region of interest: Source position is confined to a 2.8×2.1m area
- 3 microphone pairs with inter-distance of 0.2m (position unknown)



Recordings Setup

Setup:

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



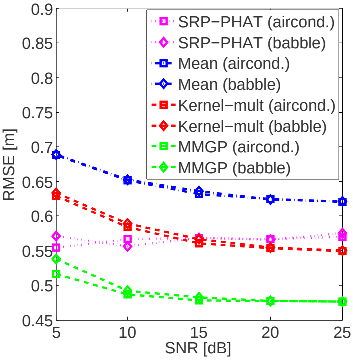
Experimental Results [Laufer-Goldshtein et al., 2016d]

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

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

  which is a global measure of smoothness for f
- Stokes' theorem links gradient and Laplacian:

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

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

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

Manifold Regularization

Graph Laplacian:

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

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

  where $\mathcal{N}_j$ is a set consisting of the d nearest-neighbours of $\mathbf{h}_j$.
- The graph Laplacian of G: $\mathbf{M} = \mathbf{S} - \mathbf{W}$, where $S_{ii} = \sum_{j=1}^{n} \mathbf{W}_{ij}$.
- Smoothness functional of G:

$$\langle \mathbf{f}_D, \mathbf{M}\mathbf{f}_D \rangle = \mathbf{f}_D^T \mathbf{M} \mathbf{f}_D$$

  where $\mathbf{f}_D = [f(\mathbf{h}_1), ..., f(\mathbf{h}_{n_D})]$.
- It can be shown: Back

$$\mathbf{f}_D^T \mathbf{M} \mathbf{f}_D = \frac{1}{2} \sum_{i,j=1}^{n_D} W_{ij} \left(f(\mathbf{h}_i) - f(\mathbf{h}_j)\right)^2$$

References

[Allen and Berkley, 1979] Allen, J. and Berkley, D. (1979).
Image method for efficiently simulating small-room acoustics.
J. Acoustical Society of America, 65(4):943–950.

[Belkin et al., 2006] Belkin, M., Niyogi, P., and Sindhwani, V. (2006).
Manifold regularization: A geometric framework for learning from labeled and unlabeled examples.
Journal of Machine Learning Research.

[Coifman and Lafon, 2006] Coifman, R. and Lafon, S. (2006).
Diffusion maps.
Appl. Comput. Harmon. Anal., 21:5–30.

[Dal-Degan and Prati, 1988] Dal-Degan, N. and Prati, C. (1988).
Acoustic noise analysis and speech enhancement techniques for mobile radio application.
Signal Processing, 15(4):43–56.

[Deleforge et al., 2015] Deleforge, A., Forbes, F., and Horaud, R. (2015).
Acoustic space learning for sound-source separation and localization on binaural manifolds.
International journal of neural systems, 25(1).

[DiBiase et al., 2001] DiBiase, J. H., Silverman, H. F., and Brandstein, M. S. (2001).
Robust localization in reverberant rooms.
In *Microphone Arrays*, pages 157–180. Springer.

[Gannot et al., 2001] Gannot, S., Burshtein, D., and Weinstein, E. (2001).
Signal enhancement using beamforming and nonstationarity with applications to speech.
IEEE Transactions on Signal Processing, 49(8):1614 –1626.

[Habets and Gannot, 2007] Habets, E. and Gannot, S. (2007).
Generating sensor signals in isotropic noise fields.
The Journal of the Acoustical Society of America, 122:3464–3470.

References (cont.)

[Jot et al., 1997] Jot, J.-M., Cerveau, L., and Warusfel, O. (1997).
Analysis and synthesis of room reverberation based on a statistical time-frequency model.
In *Audio Engineering Society Convention 103*. Audio Engineering Society.

[Knapp and Carter, 1976] Knapp, C. and Carter, G. (1976).
The generalized correlation method for estimation of time delay.
IEEE Transactions on Acoustic, Speech and Signal Processing, 24(4):320–327.

[Koldovsky et al., 2014] Koldovsky, Z., Malek, J., and Gannot, S. (2014).
Spatial source subtraction based on incomplete measurements of relative transfer function.
IEEE Transactions on Audio, Speech, and Language Processing, 23(8):1335–1347.

[Laufer-Goldshtein et al., 2015] Laufer-Goldshtein, B., Talmon, R., and Gannot, S. (2015).
Study on manifolds of acoustic responses.
In *Interntional Conference on Latent Variable Analysis and Signal Seperation (LVA/ICA)*, Liberec, Czech Republic.

[Laufer-Goldshtein et al., 2016a] Laufer-Goldshtein, B., Talmon, R., and Gannot, S. (2016a).
Manifold-based Bayesian inference for semi-supervised source localization.
In *IEEE International Conference on Audio and Acoustic Signal Processing (ICASSP)*, Shanghai, China.

[Laufer-Goldshtein et al., 2016b] Laufer-Goldshtein, B., Talmon, R., and Gannot, S. (2016b).
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.

[Laufer-Goldshtein et al., 2016c] Laufer-Goldshtein, B., Talmon, R., and Gannot, S. (2016c).
Semi-supervised sound source localization based on manifold regularization.
IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24(8):1393–1407.

[Laufer-Goldshtein et al., 2016d] Laufer-Goldshtein, B., Talmon, R., and Gannot, S. (2016d).
Semi-supervised source localization on multiple-manifolds with distributed microphones.
arXiv preprint arXiv:1610.04770.

References (cont.)

[Laufer-Goldshtein et al., 2017] Laufer-Goldshtein, B., Talmon, R., and Gannot, S. (2017).
Speaker tracking on multiple-manifolds with distributed microphones.
In *The 13th International Conference on Latent Variable Analysis and Signal Separation (LVA-ICA)*, Grenoble, France.

[Markovich-Golan and Gannot, 2015] Markovich-Golan, S. and Gannot, S. (2015).
Performance analysis of the covariance subtraction method for relative transfer function estimation and comparison to the covariance whitening method.
In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 544–548, Brisbane, Australia.

[Peterson, 1986] Peterson, P. (1986).
Simulating the response of multiple microphones to a single acoustic source in a reverberant room.
J. Acoust. Soc. Am, 76(5):1527–1529.

[Polack, 1993] Polack, J.-D. (1993).
Playing billiards in the concert hall: The mathematical foundations of geometrical room acoustics.
Applied Acoustics, 38(2):235–244.

[Schroeder, 1996] Schroeder, M. R. (1996).
The "schroeder frequency" revisited.
The Journal of the Acoustical Society of America, 99(5):3240–3241.

[Sindhwani et al., 2007] Sindhwani, V., Chu, W., and Keerthi, S. S. (2007).
Semi-supervised gaussian process classifiers.
In *IJCAI*, pages 1059–1064.

[Talmon et al., 2011] Talmon, R., Cohen, I., and Gannot, S. (2011).
Supervised source localization using diffusion kernels.
In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 245–248.

References (cont.)

[Talmon and Gannot, 2013] Talmon, R. and Gannot, S. (2013).
Relative transfer function identification on manifolds for supervised GSC beamformers.
In *21st European Signal Processing Conference (EUSIPCO)*, Marrakech, Morocco.

[Xiao et al., 2015] Xiao, X., Zhao, S., Zhong, X., Jones, D. L., Chng, E. S., and Li, H. (2015).
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.

[Yu et al., 2016] Yu, Y., Wang, W., and Han, P. (2016).
Localization based stereo speech source separation using probabilistic time-frequency masking and deep neural networks.
EURASIP Journal on Audio, Speech, and Music Processing, 2016(1):7.