SUPERVISED SOURCE LOCALIZATION USING DIFFUSION KERNELS

Ronen Talmon[1], *Israel Cohen*[1] *and Sharon Gannot*[2]

[1] Department of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel
{ronenta2@tx, icohen@ee}.technion.ac.il

[2] School of Engineering
Bar-Ilan University
Ramat-Gan, 52900, Israel
gannot@eng.biu.ac.il

ABSTRACT

Recently, we introduced a method to recover the controlling parameters of linear systems using diffusion kernels. In this paper, we apply our approach to the problem of source localization in a reverberant room using measurements from a single microphone. Prior recordings of signals from various known locations in the room are required for training and calibration. The proposed algorithm relies on a computation of a diffusion kernel with a specially-tailored distance measure. Experimental results in a real reverberant environment demonstrate accurate recovery of the source location.

***Index Terms*—** Source localization, acoustic localization, diffusion geometry, diffusion kernel, manifold learning

1. INTRODUCTION

Acoustic source localization has been a task that drew much research effort in the past several decades. In order to find the position of an acoustic source in a room, the source signal is usually picked up with a microphone array, and the relative delays between pairs of microphone signals need to be determined [1, 2, 3]. A different approach was first presented by Malioutov et al. [4], where a predefined grid of potential source positions was considered. The steering vector from each possible position was calculated and used to create an over complete representation of all possible source locations. Unfortunately, this contribution enables localization only in anechoic environments. In [5], Model and Zibulevsky extended [4] to support reverberant environments by pre-calculating the acoustic transfer functions from all the potential positions to the sensors. However, this approach suffers from high computational burden. In addition, accurate acoustic transfer functions may be analytically computed only for specific room layouts.

In this work we assume that the information on the source position may be conveyed by a *single* acoustic impulse response between the source and the microphone. Unfortunately, the acoustic response is unknown, may be analytically computed only in specific rooms, and its estimation based on a single microphone measurement is a completely blind task and hence very challenging. To overcome this difficulty, we require prior recordings of signals from various known locations in the room that are used for training and calibration. Based on the prior recordings, we propose a supervised single-channel algorithm for source localization.

Recently, we introduced in [6] a method to recover the controlling parameters of linear systems using diffusion kernels. In this paper, we apply our approach to the problem of source localization in reverberant rooms based on the measured signal in a single microphone. The proposed algorithm is based on a computation of a diffusion kernel with a specially-tailored distance measure. The kernel integrates together local estimates of the covariance matrices of the measurements into a global structure. This structure, often referred to as *manifold*, enables parametrization of the measurements. In particular, we show that the recovered parameters from each measurement represent the position coordinates of the source in the room. Experimental results in a real reverberant environment demonstrate accurate source localization.

This paper is organized as follows. In Section 2, we formulate the problem. In Section 3, we present the proposed algorithm for source localization. Finally, in Section 4, experimental results are shown, demonstrating the performance of the algorithm.

2. PROBLEM FORMULATION

An acoustic impulse response between a source and a microphone depends on several parameters: room dimensions; positions of the source and microphone; and reflection coefficients of the walls, floor and ceiling. In addition, the presence of objects in the room, e.g. furniture, and openings in the walls, such as windows and doors, affect the acoustic impulse response. Although in practice these parameters can easily be altered, in this work we assume they remain unchanged between the training and test stages. We consider a certain room and fix the position of the microphone. Thus, the remainder degree of freedom is the source position. Let $h_{\boldsymbol{\theta}}(n)$ denote an acoustic impulse response between the microphone and a source, at relative position $\boldsymbol{\theta} = [\phi, \theta, \rho]$, where ϕ and θ are the azimuth and elevation direction of arrival (DOA) angles and ρ is the distance between the source and the microphone.

For generating the training data, we pick m predefined positions of the source $\bar{\Theta} = \{\bar{\boldsymbol{\theta}}_1, \ldots, \bar{\boldsymbol{\theta}}_m\}$. From each position, we play an arbitrary stationary unknown input signal of finite length, and record the signal picked up with the microphone. The received signal is expressed as

$$\bar{y}_i(n) = h_{\bar{\boldsymbol{\theta}}_i}(n) * x_i(n) \tag{1}$$

where $x_i(n)$ and $\bar{y}_i(n)$ are the input and output signals of the room impulse response $h_{\bar{\boldsymbol{\theta}}_i}$ corresponding to the source location $\bar{\boldsymbol{\theta}}_i$. We repeat the measurement from each source location L times. However, the position of the source is slightly perturbed. Let $\{\boldsymbol{\theta}_{i_j}\}_{j=1}^L$ denote the small perturbations of $\bar{\boldsymbol{\theta}}_i$. Let $\{x_{i_j}(n), y_{i_j}(n)\}_{j=1}^L$ be the input and output signals corresponding to the repeated measurements. We assume these predefined locations and measurements are available beforehand and are used for training. It is worthwhile noting that in practice we may use short time intervals, and hence

This research was supported by the Israel Science Foundation (grant no. 1130/11).

we may only require a pseudo-stationary input signal, e.g., speech and music. However, this issue is beyond the scope of this paper and will be addressed in future work.

The input of the algorithm is a new measurement of an arbitrary unknown input signal from an unknown source position. Our goal in this work is to recover the source position given the measured signal based on the prior training information. For computational efficiency, we present the recovery of M source positions given M sequential measurements. Let $\Theta = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M\}$ denote the unknown M source positions corresponding to the new measurements. As in (1), we have

$$y_i(n) = h_{\boldsymbol{\theta}_i}(n) * x_i(n) \tag{2}$$

where $x_i(n)$ and $y_i(n)$ are input and output signals of finite length.

3. PROPOSED ALGORITHM

3.1. Features

The measured output signal heavily depends on the arbitrary random unknown input signal. Consequently, the information on the position of the source is weakly disclosed by the raw time domain measurements. To overcome this challenge, we propose to compute features that better convey the position, and are less dependent on the particular input signal. From (2), by assuming the input signal is zero-mean, the covariance function of the output signal $y_i(n)$ is given by

$$c_{y_i}(\tau) = h_{\boldsymbol{\theta}_i}(\tau) * h_{\boldsymbol{\theta}_i}(-\tau) * c_{x_i}(\tau) \tag{3}$$

where $c_{x_i}(\tau)$ and $c_{y_i}(\tau)$ denote the covariance functions of $x_i(n)$ and $y_i(n)$, respectively. We assume that the time interval is sufficiently short, so that the covariance of the pseudo-stationary input signal does not vary during the interval. Thus, the time variations of the covariance of the output signal only depend on the variations of the acoustic impulse response. Accordingly, for each measurement, we compute a feature vector consisting of D elements of the covariance. Let $\mathbf{c}_i$, $\bar{\mathbf{c}}_i$, and $\mathbf{c}_{i_j}$ denote the covariance elements of $y_i(n)$, $\bar{y}_i(n)$, and $y_{i_j}(n)$, respectively. We note that geometrically the vectors $\{\mathbf{c}_{i_j}\}_{j=1}^{L}$ are viewed as a "cloud" of points around $\bar{\mathbf{c}}_i$ in $\mathbb{R}^D$. As a consequence, they are utilized to estimate the local covariance matrix of $\bar{\mathbf{c}}_i$, i.e., $\boldsymbol{\Sigma}_i = \mathrm{Cov}(\bar{\mathbf{c}}_i)$ via

$$\hat{\boldsymbol{\Sigma}}_i = \frac{1}{L}\sum_{j=1}^{L} \mathbf{c}_{i_j}\mathbf{c}_{i_j}^T. \tag{4}$$

3.2. Training stage

We compute an affinity matrix $\mathbf{W}$ between the m training samples in $\bar{\Theta}$. As proposed in [7, 8], the matrix klth element is calculated according to

$$\mathbf{W}_{kl} = \frac{\pi}{d_{kl}} \exp\left\{-\frac{(\bar{\mathbf{c}}_k - \bar{\mathbf{c}}_l)^T[\hat{\boldsymbol{\Sigma}}_k + \hat{\boldsymbol{\Sigma}}_l]^{-1}(\bar{\mathbf{c}}_k - \bar{\mathbf{c}}_l)}{\varepsilon}\right\} \tag{5}$$

where ε is the kernel scale and d_{kl} is the following normalization factor

$$d_{kl} = \sqrt{\det\left(\mathrm{Cov}\left(\frac{\bar{\mathbf{c}}_k + \bar{\mathbf{c}}_l}{2}\right)\right)}. \tag{6}$$

It can be shown that the distance measure used in (5) approximate the Euclidean distance between the parameters [7], i.e.,

$$\left\|\bar{\boldsymbol{\theta}}_k - \bar{\boldsymbol{\theta}}_l\right\|^2 \approx (\bar{\mathbf{c}}_k - \bar{\mathbf{c}}_l)^T[\boldsymbol{\Sigma}_k + \boldsymbol{\Sigma}_l]^{-1}(\bar{\mathbf{c}}_k - \bar{\mathbf{c}}_l). \tag{7}$$

This is the key point of this work: the proposed kernel enables to capture the actual variability in terms of the source position based on the measurements.

Let $\{\lambda_j\}_{j=0}^{m-1}$ and $\{\boldsymbol{\varphi}_j\}_{j=0}^{m-1}$ be the eigenvalues and eigenvectors of the affinity matrix $\mathbf{W}$, where the eigenvalues are denoted in descending order. We note that $\lambda_0 = 1$ and the corresponding eigenvector $\boldsymbol{\varphi}_0$ is trivial [7]. There exist eigenvectors, which can be chosen as suggested in [9], that represent the data in terms of its independent parameters. For simplicity, we assume these eigenvectors correspond to the largest eigenvalues, namely, $\boldsymbol{\varphi}_1$, $\boldsymbol{\varphi}_2$, and $\boldsymbol{\varphi}_3$. In our work, these independent parameters represent the desired position coordinates of the source.

3.3. Test stage

Given a set of M new sequential measurements we compute their corresponding covariance elements $\{\mathbf{c}_i\}_{i=1}^{M}$. Let $\mathbf{A}$ be an M by m matrix computed by

$$\mathbf{A}_{kl} = \exp\left\{-\frac{(\mathbf{c}_k - \bar{\mathbf{c}}_l)^T\hat{\boldsymbol{\Sigma}}_l^{-1}(\mathbf{c}_k - \bar{\mathbf{c}}_l)}{\varepsilon}\right\}. \tag{8}$$

We note that unlike the kernel in (5) where the covariance matrices of both vectors are required, the computation of (8) involves just the available information at this point: the feature vector of the new measurement and the training data. Let $\tilde{\mathbf{A}}$ be a corresponding normalized matrix given by $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{S}^{-1/2}$, where $\mathbf{S} = \mathrm{diag}\{\mathbf{A}^T\mathbf{A}\mathbf{1}\}$ is a diagonal matrix and $\mathbf{1}$ is a vectors of ones. The normalized matrix satisfies $\mathbf{W} = \tilde{\mathbf{A}}^T\tilde{\mathbf{A}}$. Therefore, the eigenvectors of $\mathbf{W}$ of length m are the left singular vectors of $\tilde{\mathbf{A}}$ and are assumed to describe the m training measurements. The right singular vectors of $\tilde{\mathbf{A}}$ of length M are given by

$$\boldsymbol{\psi}_j = \frac{1}{\sqrt{\lambda_j}}\tilde{\mathbf{A}}\boldsymbol{\varphi}_j. \tag{9}$$

The computation of the right singular vectors via the weighted interpolation of the eigenvectors in (9) circumvents additional spectral decomposition. In addition, the right singular vectors can be viewed as the extension of the spectral representation describing the new M measurements [6].

Let Ψ be the embedding of the measurements onto the space spanned by the right singular vectors corresponding to the source position, i.e.,

$$\Psi : \mathbf{c}_i \mapsto \left[\boldsymbol{\psi}_1^{(i)}, \boldsymbol{\psi}_2^{(i)}, \boldsymbol{\psi}_3^{(i)}\right]^T. \tag{10}$$

It is shown in [6] that (10) maps the measurements into the independent parametric domain. In this case, we show in Section 4 that the map $\Psi(\mathbf{c}_i)$ indeed recovers the position of the source up to a monotonic distortion. The distortion monotonicity enables the map to organize the measurements according to the values of the source position coordinates. Thus, in order to obtain an estimate of the position, we interpolate the training positions according to distances in the embedded space. Let $\mathcal{N}_i$ consist of the k-nearest training measurements $\{\bar{\mathbf{c}}_j\}$ of $\mathbf{c}_i$ in the embedded space. Let $\{\gamma_j\}_{j=1}^{k}$ be interpolation coefficients, satisfying $\sum_{j=1}^{k}\gamma_j(\mathbf{c}_i) = 1$, given by

$$\gamma_j(\mathbf{c}_i) = \frac{\exp\left(-\|\Psi(\mathbf{c}_i) - \Psi(\bar{\mathbf{c}}_j)\|^2/\varepsilon_{\gamma_j}\right)}{\sum\limits_{\bar{\mathbf{c}}_k \in \mathcal{N}_i}\exp\left(-\|\Psi(\mathbf{c}_i) - \Psi(\bar{\mathbf{c}}_k)\|^2/\varepsilon_{\gamma_j}\right)} \tag{11}$$

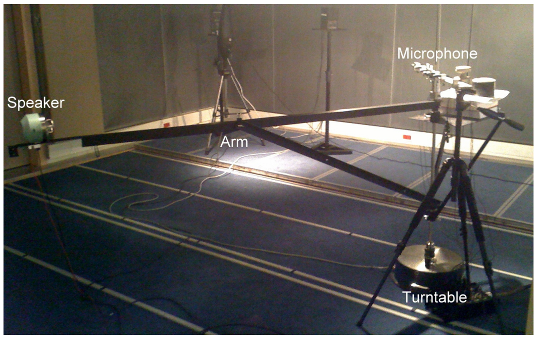

Figure 1: Recording room setup.

where ε_{γ_j} is set to the minimal distance between $\Psi(\mathbf{c}_i)$ and its nearest neighbor. Thus, an estimate of the source position is given by the following weighted sum of training locations

$$\hat{\boldsymbol{\theta}}_i = \sum_{\bar{\mathbf{c}}_j \in \mathcal{N}_i} \gamma_j(\mathbf{c}_i)\bar{\boldsymbol{\theta}}_j. \tag{12}$$

Accordingly, let $e(\mathbf{c}_i)$ be the estimation error, defined by

$$e(\mathbf{c}_i) = \|\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i\|. \tag{13}$$

3.4. One dimensional case

It can be shown that the eigenvectors are approximations of the eigenfunctions of a Laplacian operator [7]. The eigenfunctions of a one dimensional Laplacian (with Neumann boundary conditions) with a uniformly distributed parameter x on $[0,1]$ are given by

$$\varphi_n(x) = \cos(n\pi x). \tag{14}$$

Since $\boldsymbol{\psi}_j$ is similar to $\boldsymbol{\varphi}_j$, (14) implies that $\boldsymbol{\psi}_j$ represents the parameter with a *cosine* distortion. It is worthwhile noting, that the cosine function is monotonic in $[0,1]$ and therefore organizes the measurements according to the value of the parameter x.

Thus, in a special case, where we are interested just in a single coordinate, e.g. the azimuth angle ϕ, and fix the elevation angle θ and the radius ρ, we may apply the function *arccos* on the map Ψ, now comprises of one vector, in order to compensate for the distortion of the embedded points.

3.5. Scale adjustment

The scale of the kernel (5) is of key importance. As discussed in [10], setting the scale conveys a tradeoff between integration of large number of samples (large scale), and locality (small scale). Our experimental results in Section 4 demonstrate this tradeoff. An empirical approach to set the scale is to compute the root mean square error (RMSE) over the training data for various scales. The scale that yields the minimum error is set, i.e.

$$\varepsilon^* = \underset{\varepsilon}{\operatorname{argmin}} \sqrt{\frac{1}{m}\sum_{i=1}^{m} e^2(\bar{\mathbf{c}}_i)}. \tag{15}$$

In order to verify the selection of the scale, we observe the spectrum of the kernel matrix $\mathbf{W}$. In particular, we examine the location of the spectral gap, i.e., the first significant difference between consecutive eigenvalues $\{\lambda_j\}$. Ideally, the spectral gap indicates the number of degrees of freedom. Accordingly, in our general localization problem, a proper selection of the scale should yield a spectral gap after 3 (nontrivial) eigenvalues.

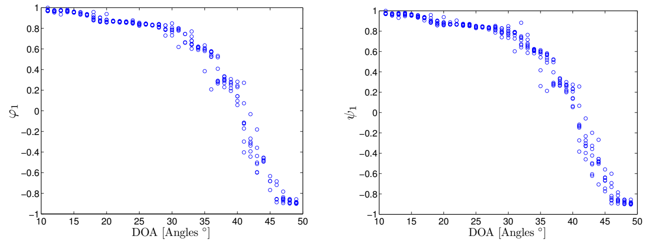

Figure 2: The embedding of the measurements as a function of the DOA azimuth angle. (a) The values of the eigenvector $\boldsymbol{\varphi}_1$. (b) The values of the extended eigenvector $\boldsymbol{\psi}_1$.

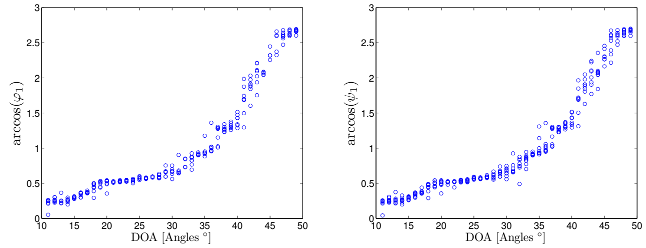

Figure 3: The embedding of the measurements after re-adjustment as a function of the DOA azimuth angle. (a) The values of $\arccos(\boldsymbol{\varphi}_1)$. (b) The values of $\arccos(\boldsymbol{\psi}_1)$.

4. EXPERIMENTAL RESULTS

In this section, we test the ability of the proposed algorithm to recover the location of an acoustic source. We conducted recordings in an acoustic room of dimensions 6 × 6 × 2.4m. The room reverberations can be controlled by 60 double-sided panels (either absorbing or reflecting) tiled over the room walls, ceiling and floor. We arranged the panels to yield moderate room reverberation time of $T_{60} = 0.3$s. Inside the room, we positioned an omni-microphone (AKG CK32) in a fixed location. A 2m long "arm" was connected to the base of the microphone, and attached to Brüel & Kajer 9640 turntable that controls the horizontal angle of the arm. A sound source (type 4227 Brül & Kjaer mouth-simulator) was located on the far-end of the arm. Thus, the turntable controlled the azimuth angle of the DOA of the sound played by the speaker with respect to the microphone. Figure 1 depicts the recording room setup.

Using the turntable we tested 60 different DOA angles with 1° spacing. In each location, 10 seconds of a zero-mean and unit-variance (neglecting the gain of the electronic system) white Gaussian noise sampled at 48 kHz was played from the mouth-simulator. The movement of the arm along the entire range of 60 angles was repeated 8 times. Consequently, we obtained 480 measurements of 10s long each, originating from 60 different angles. Due to small perturbations of the long arm, we assume that the exact location is not maintained during the entire 10s period. Thus, each measurement was divided into 10 segments of 1s each. Based on each 1s measurement we estimated $D = 100$ elements of the covariance

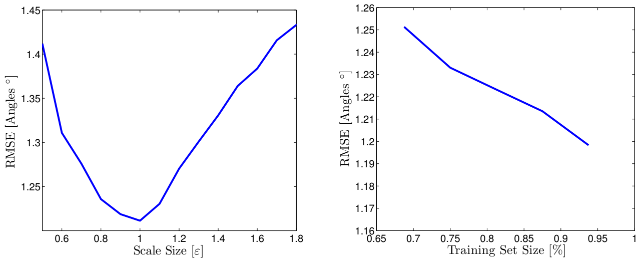

Figure 4: Localization results. The RMSE as a function of (a) the kernel scale, and (b) the test set size relative to the training set size.

function of the signal, and collected them into a feature vector. The 10 feature vectors corresponding to different 1s segments from the same DOA angle are viewed as a “cloud” of 10 points, and used to estimate the covariance matrix according to (4).

The DOA azimuth angle constitutes the sole degree of freedom in this experiment, as the rest of the room parameters, and the location of the microphone are fixed. It is worthwhile noting that we neglect the mild variations of the room impulse response caused by the movement of the arm. Thus, in this particular experiment, the map (10) is reduced to the right singular vector $\boldsymbol{\psi}_1$.

In the first experiment, we tested the ability of the proposed method to organize the recordings according to the DOA azimuth angle. We randomly chose 60 measurements for the test set, and the rest 420 were used for training. To avoid the boundary distortions imposed by the *cos* function on the embedding, we limit the observation to the range of angles $10^\circ - 50^\circ$. In Fig. 2 we show the 1-D embedding of the measurements. Figure 2(a) shows a scatter plot of the values of the eigenvector $\boldsymbol{\varphi}_1$ and (b) show a scatter plot of the values of the right singular vector $\boldsymbol{\psi}_1$, both as a function of the azimuth angle of each embedded measurement. We observe that the embedding organizes the measurements according to the azimuth angle in a monotonic order. In addition, we note that the right singular vector is similar to the eigenvector computed based solely on the training. As mentioned in Section 3, the shape of the embedding indeed corresponds with the shape of the *cos* function. Therefore, we employ the function *arccos*. The adjusted embedding is presented in Fig. 3. We now observe that the embedding via the adjusted eigenvectors maps the measurements almost linearly with respect to the DOA angle.

In the second experiment, we tested the ability of the proposed method to recover the azimuth angle. The recordings were divided into training and test sets with sizes determined by a predefined ratio. First, we randomly chose 60 measurements for testing, and the rest 420 measurements are used for training. This division with different selection of training and test sets was repeated 1000 times to yield confident results. To evaluate the localization results, we computed the RMSE of the azimuth estimate (13). Figure 4(a) shows the RMSE as a function of the kernel scale. We observe that maximum accuracy of 1.21° is obtained using $\varepsilon = 1$. In addition, it demonstrates the consideration of setting the optimal scale: smaller scale corresponds to better “spatial” resolution, whereas larger scale integrates together more points. Figure 4(b) presents the recovery accuracy based on different relative sizes of the training set with respect to the entire set. As expected, we observe that as the training set is larger, the recovery of the azimuth angle is more accurate. However, even in case the test set consists of 150 measurements out of the entire 480 (31.25%), the extended embedding is still accurate and the estimation error of the azimuth angle is small.

5. CONCLUSIONS

We have presented a supervised algorithm for source localization using a diffusion kernel. Unlike common model-based localization algorithms, the proposed algorithm entails a data-driven approach that exploits prior measurements for training and calibration. Moreover, it is based solely on single-channel recordings. Experimental results conducted in a real reverberant environment showed accurate estimation of the source direction of arrival.

For future work we intend to broaden the scope of this work. We plan to extend the algorithm for noisy environments and develop a multi-channel algorithm. In addition, it would be interesting to investigate the influence of environmental changes following the training stage, e.g. when furniture are added or when people are moving around the speaker.

6. ACKNOWLEDGMENT

The authors would like to thank Avinoam Levi for his help with the recordings, and to Udi Memshitz for designing and installing the arm.

7. REFERENCES

[1] C. Knapp and G. Carter, “The generalized correlation method for estimation of time delay,” *IEEE Trans. on Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[2] J. Chen, J. Benesty, and Y. Huang, “Time delay estimation in room acoustic environments: an overview,” *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 170–170, Jan. 2006.

[3] J. P. Dmochowski and J. Benesty, “Steered beamforming approaches for acoustic source localization,” *I. Cohen, J. Benesty, and S. Gannot (Eds.), Speech Processing in Modern Communication, Springer*, pp. 307–337, 2010.

[4] D. M. Malioutov, M. Cetin, and A. S. Willsky, “A sparse signal reconstruction perspective for source localization with sensor arrays,” *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[5] D. Model and M. Zibulevsky, “Signal reconstruction in sensor arrays using sparse representations,” *Signal Process.*, vol. 86, pp. 624–638, 2006.

[6] R. Talmon, D. Kushnir, R. R. Coifman, I. Cohen, and S. Gannot, “Parametrization of linear systems using diffusion kernels,” *submitted to IEEE Trans. Signal Process.*, 2011.

[7] A. Singer and R. Coifman, “Non-linear independent component analysis with diffusion maps,” *Appl. Comput. Harmon. Anal.*, vol. 25, pp. 226–239, 2008.

[8] D. Kushnir, A. Haddad, and R. Coifman, “Anisotropic diffusion on sub-manifolds with application to earth structure classification,” *submitted*, 2010.

[9] A. Singer, “Spectral independent component analysis,” *Appl. Comput. Harmon. Anal.*, vol. 21, pp. 128–134, 2006.

[10] M. Hein and J. Y. Audibert, “Intrinsic dimensionality estimation of submanifold in r^d,” *L. De Raedt, S. Wrobel (Eds.), Proc. 22nd Int. Conf. Mach. Learn., ACM*, pp. 289–296, 2005.