Signal Processing 108 (2015) 521–529

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Out-of-sample extension of band-limited functions on homogeneous manifolds using diffusion maps

Saman Mousazadeh*, Israel Cohen

Department of Electrical Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel

ARTICLE INFO

Article history:
Received 24 December 2013
Received in revised form
21 September 2014
Accepted 17 October 2014
Available online 27 October 2014

Keywords:
Sampling theorem
Homogenous manifolds
Band-limited functions
Laplace–Beltrami operator

ABSTRACT

In this paper, we address the problem of function extension when the available data lies on a homogeneous manifold (i.e. the domain of the function is a homogeneous manifold embedded in the Euclidean space) and the function is band-limited. We solve this problem in the general case in which the manifold is unknown. We assume that we have sufficient labeled data to reconstruct the function from labeled data. We also assume that we have enough data (at least exponential in the intrinsic dimension of the manifold) to approximate the Laplace–Beltrami operator on the manifold. The proposed method has a closed form solution and consists of matrix multiplication and inversion. As the size of data approaches infinity, the proposed method converges to the optimal solution as long as the function values are known on an appropriate sampling set. Simulation results demonstrate the advantage of the proposed method over commonly used function extension methods.

1. Introduction

Supervised learning is a machine learning task of inferring a function from labeled training data [1]. The type of data and properties of the function are application dependent. One of the most basic supervised learning problem in the field of signal processing is sampling and reconstruction of a band-limited function. When the data points are real numbers and the function is band-limited, the classical Nyquist theorem [2] states that the function can be perfectly reconstructed from its values on equally spaced points of reals, if the sampling rate is sufficiently high. The values of the function on points other than sampling points can be exactly calculated using SINC interpolator. Schoenberg [3] used cardinal splines for the reconstruction formula. There, it is shown that a band-limited function can be reconstructed from its values sampled at high enough rate (Nyquist rate) as accurately as needed using cardinal splines of sufficiently high degree. This result was further generalized to the case of nonuniform sampling by Lyubarskii and Madych [4]. More specifically, they showed that a band-limited function $f(x)$, whose Fourier transform is compactly supported between $[-\pi,\pi]$ can be completely reconstructed using spline functions, from its samples $f(x_n)$ taken at sampling points x_n, in the case when the functions $\exp(jx_n\omega)$, form a Riesz basis for $L_2([-\pi,\pi])$.

Pesenson [5] generalized the concept of band-limited functions to the case that the domain of the function is a homogeneous manifold and introduced the spectral entire functions of exponential type and Lagrangian splines on homogeneous manifolds. He also showed that on manifolds, the reconstruction of irregularly sampled spectral entire functions of exponential type (from now on band-limited functions) by splines is possible, as long as the distance between points of a sampling sequence is small enough.

Recently, using a different point of view, Coifman and Lafon [6] proposed a simple scheme, based on the Nyström

* Corresponding author.
E-mail addresses: smzadeh@tx.technion.ac.il (S. Mousazadeh), icohen@ee.technion.ac.il (I. Cohen).

http://dx.doi.org/10.1016/j.sigpro.2014.10.024
0165-1684/

method for supervised learning and extending empirical functions defined on a set X to a larger set $\overline{X}$. The extension process involves the construction of a specific family of functions termed geometric harmonics. These functions constitute a generalization of the prolate spheroidal wave functions of Slepian in the sense that they are optimally concentrated on X. Although being a powerful tool for function extension, this scheme does not make use of unlabeled data to improve the approximation of the Laplace–Beltrami operator on the manifold.

Supervised learning can be also regarded as the problem of function extension. The central dogma for studying the problem of function extension on manifolds is that the distribution of natural data is non-uniform and concentrates around low-dimensional structures. The shape (geometry) of the distribution can be exploited for efficient learning. As a justification for manifold assumption of natural data, see Jansen and Niyogi [7] for speech signals and Donoho and Grimes [8] for images. Geometrically derived methods have been used in applications such as image clustering [9,10], image completion [11], speech enhancement in presence of transient noise [12], voice activity detection in presence of transient noise [13], linear and nonlinear independent component analysis [14,15], parametrization of linear systems [16], and single channel source localization [17].

Zhu et al. [18] introduced an approach for supervised learning which is based on a Gaussian random field model. Labeled and unlabeled data were represented as vertices in a weighted graph, with edge weights encoding the similarity between instances. The learning problem was then formulated in terms of a Gaussian random field on this graph, where the mean of the field was characterized in terms of harmonic functions, and was efficiently obtained using matrix methods or belief propagation. In [19], it is shown that this method becomes ill posed as the number of unlabeled points tends to infinity. This observation was the motivation for Zhou and Belkin [20] to address the semi-supervised learning problem and propose a solution by using regularization based on an iterated Laplacian, which is equivalent to a higher order Sobolev semi-norm. Their proposed solution can alternatively be viewed as a generalization of the thin plate spline to an unknown submanifold in high dimensions.

In most practical applications, besides the function, the data manifold is also unknown and just some labeled data (the points that the value of function is known on them, i.e. sampling points) and unlabeled data (the points that the value of function on them must be determined, i.e. interpolation points) are available. This means that the sampling theorem on manifolds [5] cannot be utilized directly to learn and extend the function to unlabeled data because in [5], the manifold is assumed to be known.

In order to be able to use a sampling theorem on manifolds, one needs to completely know the manifold. This means that the Laplace–Beltrami operator on the manifold must be known and can be computed for every function. Many manifold learning algorithms have been introduced during the last decade, among them one can name isomap [21], Locally-linear embedding (LLE) [22], Laplacian eigenmaps [23] and diffusion maps [24]. Diffusion maps leverage the relationship between heat diffusion and a random walk on a graph. The heat diffusion on manifold, is the diffusion process whose infinitesimal generator is the Laplace–Beltrami operator. In [24], an analogy is drawn between the diffusion operator on a manifold and a Markov transition matrix operating on functions defined on the graph whose nodes were sampled from the manifold. It is also shown that one can approximate the Laplace–Beltrami operator using appropriately normalized Markov transition matrix.

In this paper, we propose a novel technique for supervised learning when the data is assumed to be located on a manifold. More specifically, we use diffusion maps as a tool for manifold learning and approximating the Laplace–Beltrami operator on a manifold. Next, we use sampling theorem of band-limited functions on manifolds [5] to extend the function onto the interpolation points. The solution coincides with the method proposed in [20], hence gives another justification for the method presented in [20]. This paper is organized as follows. In Section 2, we formulate the problem and introduce our function extension algorithm. In Section 3, we evaluate the performance of our method and compare it to several available function extension methods. We also discuss some applications of the proposed method. We conclude the paper in Section 4.

2. Problem formulation

Let $\mathcal{M}$ be a C^∞-homogeneous manifold and Δ be the Laplace–Beltrami operator in the corresponding Hilbert space $L^2(\mathcal{M})$. We say that a function $f(\cdot)$ from $L^2(\mathcal{M})$ is ω_0-band-limited if the function satisfies the Bernstein inequality [5]:

$$\|\Delta^{k/2} f\| \le \omega_0^k \|f\| \tag{1}$$

for every natural even k, where $\|f\|$ denotes $L^2(\mathcal{M})$ norm. Using Parseval's theorem, it can be easily verified that in the special case where $\mathcal{M}=\mathbb{R}$, this definition is equivalent to the definition of band-limited functions (i.e. the Fourier transform is compactly supported in $[-\omega_0,\omega_0]$).

A set of points $Z_\lambda=\{\boldsymbol{x}_\gamma\}$, is called a sampling sequence if

(a) $\inf_{\gamma\neq\mu} \operatorname{dist}(\boldsymbol{x}_\gamma,\boldsymbol{x}_\mu)>0$,
(b) Balls $B(\boldsymbol{x}_\gamma,\lambda)$ form a cover of $\mathcal{M}$, and
(c) $\lambda<(c_0\omega_0)^{-1}$,

where c_0 is a manifold-dependent constant. In the case $\mathcal{M}=\mathbb{R}$, the last condition becomes the Nyquist sampling condition if the sampling is uniform. It can be shown [5] that any ω_0-band-limited function can be exactly reconstructed from its samples as long as the value of the function is known on a sampling sequence.

In [5], it is shown that given an ω_0-band-limited function $f(\cdot)$ on a d-dimensional manifold $\mathcal{M}$, $\epsilon>0$ and a sampling sequence Z_λ, there exists a function $\hat{f}^k$ such that

$$\|f-\hat{f}^k\|<\epsilon;\quad k=2^l d,\quad l\in\mathbb{N} \tag{2}$$

for a sufficiently large l. The function $\hat{f}^k$ is the solution of the following optimization problem:

$$\hat{f}^k=\arg\min_u \|\Delta^{k/2}u\|$$

s.t.

$$\hat{f}^k(\boldsymbol{x}_\gamma) = f(\boldsymbol{x}_\gamma); \quad \forall \boldsymbol{x}_\gamma \in Z_\lambda. \tag{3}$$

In [5], it has been shown that the above optimization problem has a unique solution for each k, but the author did not supply any algorithm for solving it in the general case. The main contribution of this paper is to present a method for solving the above optimization problem in the case where no information about the manifold is available (i.e. the Laplace–Beltrami operator is unknown). We assume that we have a valid sampling sequence (i.e. a sequence satisfying the conditions (a) through (c) above) so that we can exact reconstruction is possible. We also assume that we have enough (theoretically infinite but at least exponential in the intrinsic dimension of the manifold) unlabelled data the approximation of the Laplace–Beltrami is sufficiently accurate. This issue will be discussed later in more details.

Diffusion maps [24] is a powerful tool for nonlinear dimensionality reduction. Besides being a useful means for dimensionality reduction, it can be used for manifold learning. Using anisotropic kernel for constructing diffusion maps, Nadler et al. [25], described a random walk construction that in the limit of infinite data recovers the Laplace–Beltrami (heat) operator on the manifold on which the data resides, regardless of the possibly non-uniform sampling of points on it. This normalization is therefore best suited for learning the geometry of the data-set, as it separates the geometry of the manifold from the statistics on it. In what follows we briefly overview the method and introduce a novel technique for solving the optimization problem in (3).

Let $X = \{\boldsymbol{x}_j\}_{j=1}^{J}$ be a finite data set sampled from a homogeneous manifold $\mathcal{M} \subset \mathbb{R}^P$ randomly sampled from some arbitrary probability distribution. Suppose that we are given a kernel $k: X \times X \to \mathbb{R}$ that is symmetric (i.e. $k(\boldsymbol{x},\boldsymbol{y}) = k(\boldsymbol{y},\boldsymbol{x})$) and is positive (i.e. $k(\boldsymbol{x},\boldsymbol{y}) \geq 0$). This kernel represents some notion of affinity or similarity between points of X. This kernel describes the relationship between pairs of points in the set X and in this sense, one can think of the data points as being the nodes of a symmetric graph whose weight function is specified by k. Without loss of generality, for a fixed value of ϵ (a meta-parameter of the algorithm), we define an isotropic diffusion kernel:

$$k_\epsilon(\boldsymbol{x},\boldsymbol{y}) = \exp\left(\frac{-\|\boldsymbol{x}-\boldsymbol{y}\|^2}{\epsilon}\right). \tag{4}$$

We construct an $J \times J$ similarity matrix $\boldsymbol{K}$ such that

$$\boldsymbol{K}_{i,j} = k_\epsilon(\boldsymbol{x}_i, \boldsymbol{x}_j), \tag{5}$$

where $\boldsymbol{K}_{i,j}$ is the (i,j)-th element of matrix $\boldsymbol{K}$. Next, we normalize the similarity matrix as follows:

$$\boldsymbol{W} = \boldsymbol{D}^{-1}\boldsymbol{K}\boldsymbol{D}^{-1}, \tag{6}$$

where $\boldsymbol{D}$ is a diagonal matrix whose (i,i)-th element equals to the sum of the i-th row (or equivalently column) of $\boldsymbol{K}$ (i.e. $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{K}\boldsymbol{1})$ where $\boldsymbol{1}$ is the column vector of ones). The Laplacian matrix $\boldsymbol{L}$ is then obtained by

$$\boldsymbol{P} = \boldsymbol{D}_W^{-1}\boldsymbol{W}, \tag{7}$$

$$\boldsymbol{L} = \frac{\boldsymbol{I}-\boldsymbol{P}}{\epsilon} \tag{8}$$

where $\boldsymbol{D}_W = \mathrm{diag}(\boldsymbol{W}\boldsymbol{1})$. It can be shown that both $\boldsymbol{L}$ and $\boldsymbol{L}^T$ (where $(\cdot)^T$ denotes the transpose of a matrix or a vector) converge to the Laplace–Beltrami operator when the number of data, J, approaches infinity and ϵ approaches zero [24,25]. Since our approximation of the Laplace–Beltrami operator needs to be symmetric, we choose

$$\boldsymbol{L}_s = \frac{\boldsymbol{L}+\boldsymbol{L}^T}{2}, \tag{9}$$

as a discrete approximation of the Laplace–Beltrami operator.

Let $X_s = \{\boldsymbol{x}_i^s\}_{i=1}^{M}$ be a set consisting of a labelled or training data (i.e. a sampling sequence) which satisfying the conditions (a) through (c), and let $\boldsymbol{f}^s$ be a column vector such that $\boldsymbol{f}_i^s = f(\boldsymbol{x}_i^s)$ where $\boldsymbol{f}_i^s$ is the i-th element of vector $\boldsymbol{f}^s$. Let $X_I = \{\boldsymbol{x}_i^I\}_{i=1}^{N}$ be a set consisting of unlabeled or testing data (i.e. an interpolation sequence). Let $\boldsymbol{f}^I$ be a column vector such that $\boldsymbol{f}_i^I = f(\boldsymbol{x}_i^I)$ where $\boldsymbol{f}_i^I$ is the i-th element of vector $\boldsymbol{f}^I$. The goal is to determine $\boldsymbol{f}^I$ given the value of a band-limited function $f(\cdot)$, on a sampling sequence (i.e. given vector $\boldsymbol{f}^s$). We assume that $N+M$ is sufficiently large such that the approximation of the Laplace–Beltrami operator on the manifold is accurate enough (at least exponential in the intrinsic dimension of the manifold). Note that here we have two independent sampling procedures. In order to learn the function on the manifold we need labeled data which means sampling the function defined on the manifold such that the conditions (a) through (c) are satisfied. In order to learn the manifold (i.e. approximating of the Laplace–Beltrami operator on the manifold), we need samples from the manifold, to learn the Laplace–Beltrami operator on the manifold. The diffusion maps use all the available data sampled from the manifold to find the approximation of Laplace–Beltrami operator, whether or not they are labeled.

Let $\boldsymbol{X}$ be a $(M+N) \times P$ matrix where each row represents a single data point. Without loss of generality we assume that the first M rows represent the sampling sequence (i.e. $\boldsymbol{X}_{i,:} = \boldsymbol{x}_i^s$ for $i = 1, 2, \ldots, M$, where $\boldsymbol{X}_{i,:}$ is the i-th row of matrix $\boldsymbol{X}$) and the next N rows consist of interpolation points (i.e. $\boldsymbol{X}_{i+M,:} = \boldsymbol{x}_i^I$ for $i = 1, 2, \ldots, N$). Let $\boldsymbol{f}$ be a column concatenation of $\boldsymbol{f}^s$ and $\boldsymbol{f}^I$ i.e.

$$\boldsymbol{f} = \begin{bmatrix} \boldsymbol{f}^s \\ \boldsymbol{f}^I \end{bmatrix} \tag{10}$$

Let K be the $(M+N) \times (M+N)$ similarity matrix computed by

$$\boldsymbol{K}_{i,j} = k_\epsilon(\boldsymbol{X}_{i,:}, \boldsymbol{X}_{j,:}), \tag{11}$$

and let $\boldsymbol{L}_s$ be a discrete approximation of the Laplace–Beltrami operator computed using (6) through (9). Using $\boldsymbol{L}_s$, discrete approximation of the optimization problem in (3) is obtained as follows:

$$\hat{\boldsymbol{f}}^k = \arg\min_{\boldsymbol{f}} \|\boldsymbol{L}_s^{k/2}\boldsymbol{f}\|$$

s.t.

$$\hat{\boldsymbol{f}}_i^k = \boldsymbol{f}_i^s; \quad I = 1, 2, \ldots, M, \tag{12}$$

where $\hat{\boldsymbol{f}}_i^k$ and $\boldsymbol{f}_i^s$ are the i-th elements of $\hat{\boldsymbol{f}}^k$ and $\boldsymbol{f}^s$, respectively. The above optimization problem can be equivalently written as follows:

$$\hat{\boldsymbol{f}}^k = \arg\min_{\boldsymbol{f}} \boldsymbol{f}^T \boldsymbol{L}_s^k \boldsymbol{f}$$

s.t.

$$\boldsymbol{A}\hat{\boldsymbol{f}}^k = \boldsymbol{f}^s, \tag{13}$$

where $\boldsymbol{A} = [\boldsymbol{I}_{M\times M} : \boldsymbol{0}_{M\times N}]$ and $\boldsymbol{I}_{M\times M}$ and $\boldsymbol{0}_{M\times N}$ are the identity and zero matrices of appropriate sizes, respectively. This is a simple quadratic programming problem with equality constraints and can be solved easily using range space or null space approaches [26]. Using the null space method, the solution of the optimization problem in (13) is given by

$$\hat{\boldsymbol{f}}^k = \boldsymbol{A}\boldsymbol{f}_s - \boldsymbol{Z}\left(\boldsymbol{Z}^T\boldsymbol{L}_s^k\boldsymbol{Z}\right)^{-1}\left(\boldsymbol{Z}^T\boldsymbol{L}_s^k\boldsymbol{A}\boldsymbol{f}_s\right), \tag{14}$$

where $\boldsymbol{Z} \in \mathbb{R}^{(N+M)\times(N)}$ is the matrix whose columns span the null space of $\boldsymbol{A}$ and is given by

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{0}_{M\times N} \\ \boldsymbol{I}_{N\times N} \end{bmatrix}. \tag{15}$$

The overall function extension algorithm is summarized in Table 1.

The last important issues which must be addressed here are the determination of the power of Laplacian (i.e. k), the kernel width ϵ and the effect of number of samples on the performance of the proposed methods. As is discussed in [5], if the Laplace–Beltrami operator is fully known, the error in function extension decreases by increasing k. Hence, in the proposed method, before k becomes too large, causing numerical issues, almost always larger k performs better. From the Sobolev embedding theorem, increasing k restricts solution space to be a smoother space, and from kernel point of view, increasing k corresponds to a better density adaptive kernel [20]. It can be shown that, in continuous case, the $L^2(\mathcal{M})$ norm of the error in function reconstruction decrease exponentially as k grows [5]. In practice, k can be chosen by cross-validation method.

The convergence to the continuous diffusion operator can be utilized for properly choosing the kernel width ϵ. In [27] and [28], it was proposed to automatically set the scale by examining a logarithmic scale of the sum of the kernel weights, without computing the spectral decomposition of the transition matrix. More specifically, using the fact that the logarithmic plot of $\sum_{i,j} k_\epsilon(x_i, x_j)$ with respect to $\log\epsilon$ cannot be linear for all ϵ's, they suggested to choose the kernel width ϵ from that linear region. See [27,28] and [12] for further discussion.

The last issue to be discussed here is the effect of number of samples on the performance of the proposed methods. Several papers provide rigorous estimates for the accuracy of the approximation of Laplace–Beltrami operator. For instance, it is shown in [24] that with high probability the error in approximation of Laplace–Beltrami operator using finite data is of the order $\mathcal{O}\left(J^{-1/2}\epsilon^{-d/4-1/2}\right)$ where J is the number of data points used for approximating the Laplace–Beltrami operator, d is the intrinsic dimension of the manifold and ϵ is the kernel width.

Table 1
Proposed algorithm for extension of band-limited functions on homogenous manifolds using diffusion maps.

Input

$X_s = \{\boldsymbol{x}_i^s\}_{i=1}^M$: The set containing a sampling sequence

$\boldsymbol{f}^s$: A column vector such that $\boldsymbol{f}_i^s = f(\boldsymbol{x}_i^s)$.

$X_I = \{\boldsymbol{x}_i^I\}_{i=1}^N$: The set containing the interpolation sequence.

$k_\epsilon(\cdot,\cdot)$: A kernel represents local similarity between points.

(1) Construct $X = [X_s; X_I]$ by column concatenation of sampling sequence and interpolation points.

(2) Construct the similarity matrix $\boldsymbol{K}$ such that $\boldsymbol{K}_{i,j} = k_\epsilon(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

(3) Let

$$\boldsymbol{W} = \boldsymbol{D}^{-1}\boldsymbol{K}\boldsymbol{D}^{-1},$$
$$\boldsymbol{P} = \boldsymbol{D}_W^{-1}\boldsymbol{W},$$
$$\boldsymbol{L} = \frac{\boldsymbol{I}-\boldsymbol{P}}{\epsilon},$$
$$\boldsymbol{L}_s = \frac{\boldsymbol{L}+\boldsymbol{L}^T}{2},$$

where $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{K}\boldsymbol{1})$ and $\boldsymbol{D}_W = \mathrm{diag}(\boldsymbol{W}\boldsymbol{1})$.

(4) Let

$$\hat{\boldsymbol{f}}^k = \boldsymbol{A}\boldsymbol{f}_s - \boldsymbol{Z}\left(\boldsymbol{Z}^T\boldsymbol{L}_s^k\boldsymbol{Z}\right)^{-1}\left(\boldsymbol{Z}^T\boldsymbol{L}_s^k\boldsymbol{A}\boldsymbol{f}_s\right),$$

where $\hat{\boldsymbol{f}}^k$ is the estimate of the function on all data.

Output

$\hat{\boldsymbol{f}}^I$: A column vector such that $\hat{\boldsymbol{f}}_i^I = \hat{f}(\boldsymbol{x}_i^I)$.

3. Applications, simulation results and performance evaluation

In this section, we evaluate the performance of the proposed method using simulations and compare our method to several available methods. In all simulations, the kernel width (i.e, ϵ) in all kernel based methods is chosen such that the normalized root mean square error (NRMSE) (to be defined shortly in (16)) is minimized. Although computation of NRMSE needs the knowledge of the actual function (which is the quantity we try to estimate), we use this method for all the methods to make our comparison fair. Choosing an appropriate kernel width is a critical issue in most of kernel methods (like the proposed method or geometric harmonics), but dealing with this issue is beyond the scope of this paper.

In the first example, we try to reconstruct a band-limited function defined on an interval of reals from uniformly sampled points. Note that an interval of reals is not a homogeneous manifold, but neglecting the effect of the end points it is a homogeneous manifold. We assume that the sampling rate is more than the Nyquist rate so exact reconstruction is possible. In this case, the optimal reconstruction scheme is a SINC interpolator. We try to reconstruct

$$f(x) = \sin(2\pi x) + \sin\left(4\pi x + \frac{2\pi}{5}\right) + \sin\left(6\pi x + \frac{\pi}{3}\right) + \cos(8\pi x)$$

from its samples. The sampling frequency was set to 10 Hz (a little bit more than the Nyquist rate which is 8 Hz). We try to reconstruct the function in interval $[0,10]$. We sample the manifold (i.e. the interval $[0,10]$) uniformly with frequency 100 Hz. The results of simulation are depicted in Fig. 1. The normalized root mean square error (NRMSE) in extension is computed by

$$\mathrm{NRMSE}=\sqrt{\frac{\sum_{\boldsymbol{x}\in X}|f(\boldsymbol{x})-\hat{f}(\boldsymbol{x})|^2}{\sum_{\boldsymbol{x}\in X}|f(\boldsymbol{x})|^2}} \tag{16}$$

as a function of k, and is depicted in Fig. 1. It is obvious that for a sufficiently large k, the MSE of the proposed method is lower than that of geometric harmonics. It is also apparent that the proposed method converges to the optimal solution (i.e. SINC interpolation) as k increases. Although theoretically increasing k improves the performance of the proposed function extension algorithm, it must be taken into account that increasing k while the number of available data (i.e. $N+M$) is constant, does not necessarily improve the overall performance (i.e. decreasing MSE in function extension; see the following simulations for an example). This is because of the fact that the estimation of Laplace–Beltrami operator near the boundaries of the manifold and regions with high curvature is not very accurate. Increasing k causes intense error in extension scheme because of error in estimating the Laplace–Beltrami operator near the boundaries and regions with high curvature.

In the second simulation, we try to reconstruct the same function as in the first simulation from nonuniform sampling points. In this case, a band-limited function defined on reals whose Fourier transform is compactly supported between $[-W,W]$ can be completely reconstructed using spline functions from its samples $f(x_n)$ taken at sampling points x_n, in the case when the functions $\exp(jx_n\omega)$ form a Riesz basis for $L_2([-W,W])$. Without loss of generality, suppose that $\{x_n; n\in\mathbb{Z}\}$ is an increasing sequence. It can be shown that, the sequence $\exp(jx_n\omega)$ forms a Riesz basis for $L_2([-W,W])$, if $x_n=(n\pi/W)+r_n$ where $|r_n|<\pi/4W$ [29]. In this simulation, the sampling on the manifold is exactly the same as in the previous simulation (i.e. uniform samples on interval $[0,10]$ with frequency 100 Hz). Samples of the function are taken at the same points as in the previous simulation with random uniform displacement of maximum 0.025 (i.e. $x_n=(n/10)+r_n$ where r_n is uniformly distributed in $[-0.025,0.025]$). The simulation results are depicted in Fig. 2. NRMSE as a function of k is depicted in Fig. 2. It is obvious that for k sufficiently large, the NRMSE achieved by the proposed method is lower than that obtained by geometric harmonics.

In the third simulation, we try to reconstruct the same function as in the second simulation from nonuniform sampling points while the domain is the unit circle in a two dimensional Euclidean space which is a homogeneous manifold. We choose 1500 points from the interval $[-\pi,\pi]$ sampled from normal distribution of zero mean and variance π. Then we embed them to the unit circle in a two dimensional Euclidean space. The function $f(x,y)$ is then defined as

$$f(x,y)=\sin(2\pi x)+\sin\left(4\pi x+\frac{2\pi}{5}\right)+\sin\left(6\pi x+\frac{\pi}{3}\right)+\cos(8\pi x). \tag{17}$$

The simulation results are depicted in Fig. 3. NRMSE as a function of k is depicted in Fig. 3. It is obvious that for k large enough, the NRMSE achieved by the proposed method is comparable to that obtained by geometric harmonics and much better than the method proposed in [18].

In the following example, we recover the parameters of an autoregressive-moving average (ARMA) system. Consider the following ARMA process of orders (p,q):

$$y(t)-\sum_{\ell=1}^{p}\boldsymbol{a}_\ell y(t-\ell)=x(t)+\sum_{i=1}^{q}\boldsymbol{b}_i x(t-i) \tag{18}$$

where $x(t)$ is a zero-mean white noise with variance σ_x^2, and $\{\boldsymbol{a}_\ell\}$ and $\{\boldsymbol{b}_i\}$ are the AR and the MA coefficients, respectively. Such an ARMA process is commonly used in many signal processing applications. ARMA model is appropriate when a system is a function of a series of unobserved shocks (the MA part) as well as its own behaviour. For example, stock prices may be shocked by fundamental information as well as exhibiting technical

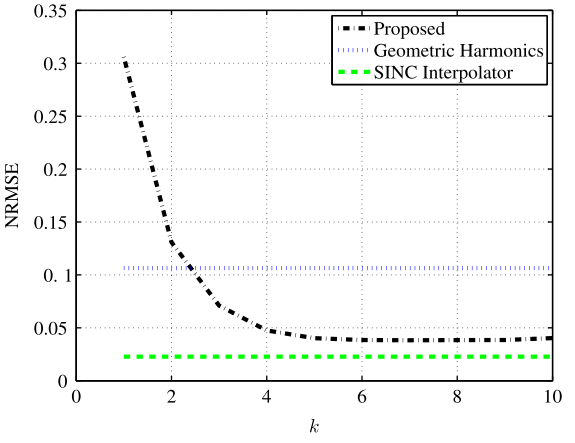

Fig. 1. Normalized root mean square error (NRMSE) in function extension for different methods as a function of the parameter k.

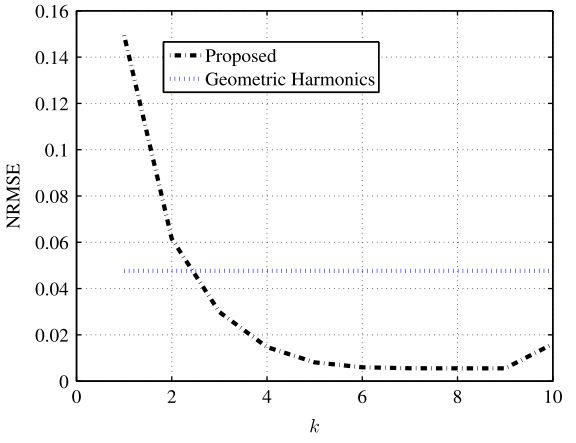

Fig. 2. Normalized root mean square error (NRMSE) in function extension for different methods as a function of the parameter k.

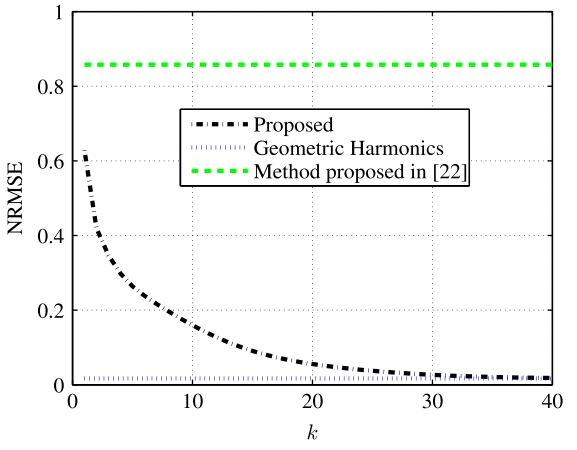

Fig. 3. Normalized root mean square error (NRMSE) in function extension for different methods as a function of the parameter k.

trending and mean-reversion effects due to market participants.

An ARMA process can be viewed as a white noise filtered by a linear system, where the corresponding transfer function is given by

$$H_{\boldsymbol{a},\boldsymbol{b}}(\omega)=\frac{1+\sum_{i=1}^{q}\boldsymbol{b}_i e^{-j\omega i}}{1-\sum_{\ell=1}^{p}\boldsymbol{a}_\ell e^{-j\omega\ell}}. \tag{19}$$

Let $S_{\boldsymbol{a},\boldsymbol{b}}(\omega)$ be the power spectral density (PSD) of the ARMA process given by

$$\begin{aligned} S_{\boldsymbol{a},\boldsymbol{b}}(\omega)&=\sigma_x^2|H_{\boldsymbol{a},\boldsymbol{b}}(\omega)|^2\\ &=\sigma_x^2\frac{|1+\sum_{i=1}^{q}\boldsymbol{b}_i e^{-j\omega i}|^2}{|1-\sum_{\ell=1}^{p}\boldsymbol{a}_\ell e^{-j\omega\ell}|^2}. \end{aligned} \tag{20}$$

We observe in last equation that the PSD depends only on the ARMA parameters (i.e. controlling parameters) when σ_x^2 assumed to be known. Consequently, the variations of the controlling parameters are conveyed by the PSD. Now, from (20), we can express the covariance function of the output signal as

$$c_{y_{\boldsymbol{a},\boldsymbol{b}}}(\tau)=\Im^{-1}\{S_{\boldsymbol{a},\boldsymbol{b}}(\omega)\}, \tag{21}$$

where $\Im^{-1}\{\cdot\}$ denotes the inverse Fourier transform. Hence, we can conclude that the PSD and consequently the covariance function of a specific output signal is a high dimensional data lying in a low dimensional manifold of dimension $p+q$ embedded in a high dimensional space.

Now if we look at each parameters of the model as function of the covariance function, our task is to extend this function from a given sampling sequence to an interpolating sequence. In this example, we examine the ability of the proposed algorithm to recover the parameters of ARMA$(1,1)$ processes. The parameters of the ARMA processes (i.e. $\boldsymbol{a}_1$ and $\boldsymbol{b}_1$) are uniformly sampled from the rectangle $[0,0.5\,0.3,1]$ (i.e. $0\le\boldsymbol{a}_1\le 0.3$ and $0.5\le\boldsymbol{b}_1\le 1$). The process noise used in this simulation is chosen to be independent identically distribute white Gaussian noise with unity variance. We generate 1000 ARMA processes with different parameters and compute 20 correlation coefficients. From these 1000, 20-dimensional correlation coefficients, we uniformly (with respect to the sampling on parameter space) chose 16 of them for training sequence and utilize the proposed algorithm to extended the function on the rest of the points. We use a Gaussian kernel for similarity computation with scale parameter set to $\epsilon=0.5$, i.e.

$$k_\epsilon\left(c_{y_{\boldsymbol{a}_1^i,\boldsymbol{b}_1^i}}(\tau),c_{y_{\boldsymbol{a}_1^j,\boldsymbol{b}_1^j}}(\tau)\right)=\exp\left(-\frac{\left\|c_{y_{\boldsymbol{a}_1^i,\boldsymbol{b}_1^i}}(\tau)-c_{y_{\boldsymbol{a}_1^j,\boldsymbol{b}_1^j}}(\tau)\right\|^2}{0.5}\right). \tag{22}$$

The results of the simulation are depicted in Figs. 4 through 5. We compare our results to those of iterative search method widely used in estimating the parameters of the ARMA model [30]. An iterative search algorithm minimizes a robustified quadratic prediction error criterion. The iterations are terminated either when the maximum number of iterations is reached, or when the expected improvement is less than the specified tolerance, or when a lower value of the criterion cannot be found. We have also compared our results with those of KNN-search and the method proposed in [18].

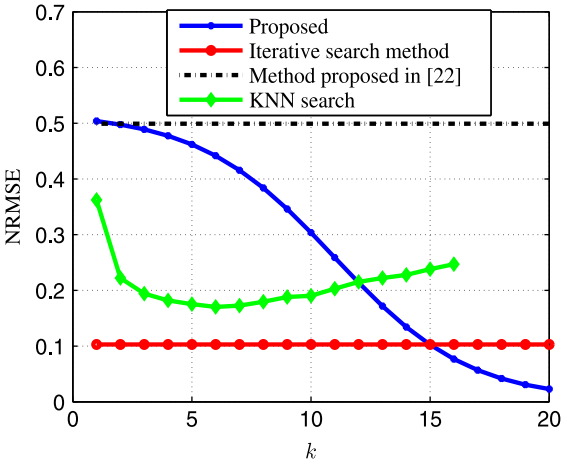

Fig. 4. Normalized root mean square error (NRMSE) in estimating the parameters of the AR part for different methods as a function of the parameter k.

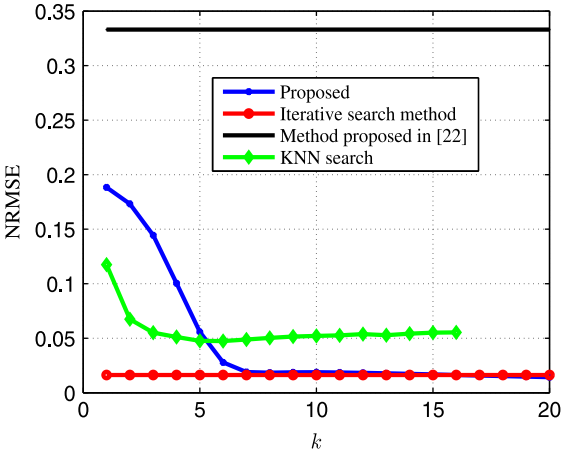

Fig. 5. Normalized root mean square error (NRMSE) in estimating the parameters of the MA part for different methods as a function of the parameter k.

In Figs. 4 and 5 the x-axis regards to the power of Laplacian for the proposed method or number of nearest neighbor in KNN-search method. As it can be seen from these Figs. 4 and 5, the proposed method outperforms the competing methods for large enough k in estimating the parameters of AR part while have better or approximately the same performance as competing method in estimating the parameters of MA part.

In our last experiment, we examine the ability of the proposed method to recover the controlling parameters of acoustic channels, simulated using the image method. The propagation of a sound wave within an enclosure can be considered linear if the medium is homogeneous. Hence, the acoustic channel from a source to a microphone is obtained by solving the wave equation. However, this solution is difficult to be expressed analytically; hence, in practice usually some approximation is utilized. The image method, presented by Allen and Berkley [31] is one of the most common methods for this task. In this method, the acoustic channel between a source and a sensor in a rectangular room is approximated by a linear time invariant (LTI) finite impulse response (FIR) system.

In order to approximate the acoustic channels in typical rooms, Habets [32] has provided a software which simulates the acoustic channel between a source and a sensor using image method which has 12 controlling parameter. These controlling parameter consist of $\boldsymbol{\beta}=[\beta_{x_1},\beta_{x_2},\beta_{y_1}$ $\beta_{y_2},\beta_{z_1},\beta_{z_2}]$ the reflection coefficients of the six walls, $\boldsymbol{r}_s=[x_s,y_s,z_s]$ and $\boldsymbol{r}_m=[x_m,y_m,z_m]$ the source and the microphone location, respectively. Typical impulse responses consist of thousands of taps. In other words, each impulse response can be expressed as a vector in a high-dimensional space. However, as we have discussed earlier, the acoustic channel between a source and a microphone inside a rectangular room is controlled by a set of 12 controlling parameter. Hence, we can consider each impulse response as a point from a low dimensional manifold embedded in a high dimensional space. Now if we regard each controlling parameter as a function defined on the impulse response, and we know the value of these parameters on a sampling sequence, we can use the proposed method to extend these function on unlabeled data (i.e. the impulse responses whose controlling parameter is unknown) in order to estimate these parameters.

In this experiment, we recover the reflection coefficients of two walls. We generate 20 training channels, where β_{x_1} is sampled uniformly from the interval [0.150.55] and we set $\beta_{x_2}=0.7-\beta_{x_1}$. The other four coefficients are set to 0.5. Then we simulate a room of size [4, 5, 3] meter. We place a microphone at $\boldsymbol{r}_m=[2,1.5,2]$, and a source at $\boldsymbol{r}_s=[2,3.5,2]$, distant 2 meters from the microphone. We then generate the acoustic impulse response of length 4096 samples and compute 30 correlation coefficients of each impulse response.

The test set is exactly obtained as in training set where for β_{x_1} we sampled 180 points uniformly from the interval [0.150.55] and we set $\beta_{x_2}=0.7-\beta_{x_1}$. All other parameters are set as in training procedure. We used a Gaussian kernel with kernel width $\epsilon=1.0710\times10^{-18}$.

The results of the simulation are depicted in Figs. 6 through 9. We compare our results to those of KNN-search.

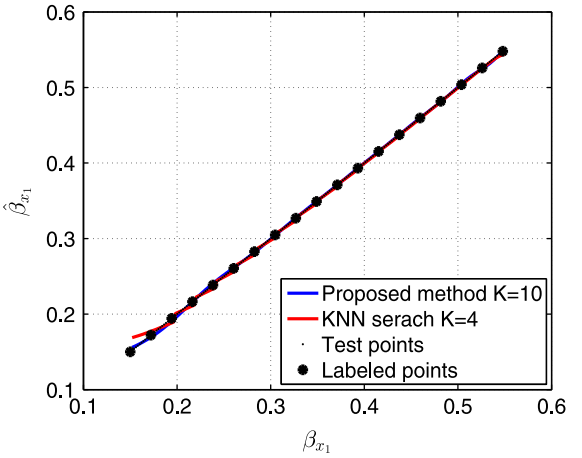

Fig. 6. Estimated reflection coefficient β_{x_1} versus the true value for different methods.

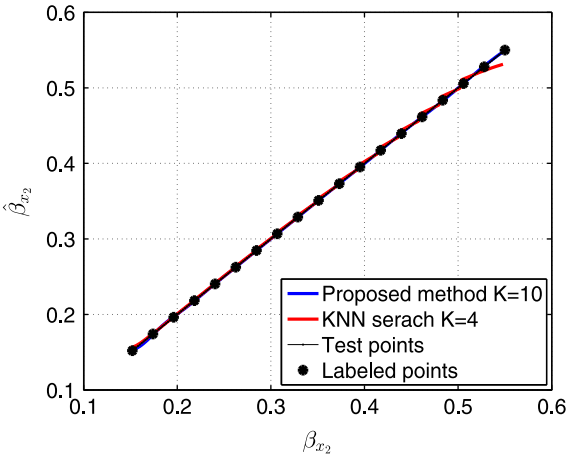

Fig. 7. Estimated reflection coefficient β_{x_2} versus the true value for different methods.

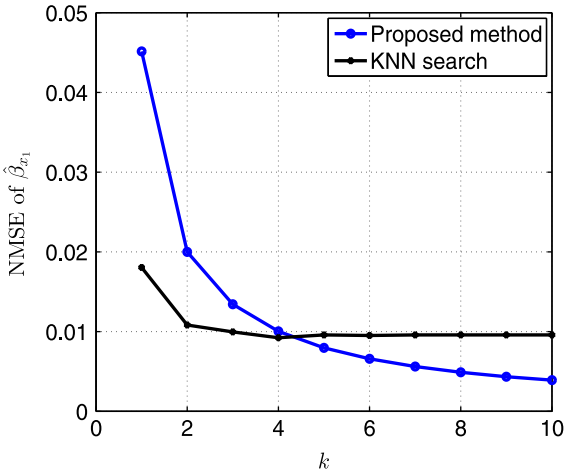

Fig. 8. Normalized root mean square error (NRMSE) in estimating the reflection coefficient β_{x_1} for different methods as a function of the parameter k.

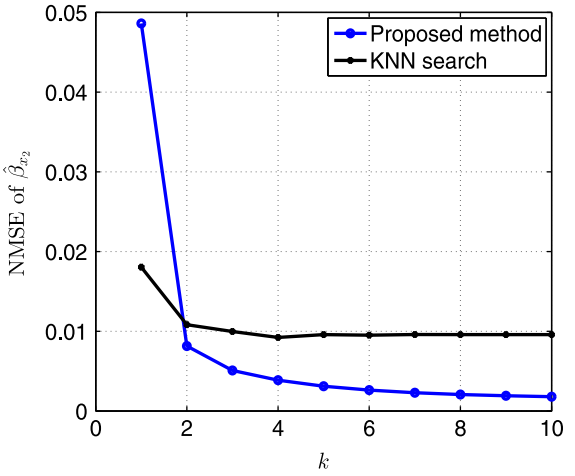

Fig. 9. Normalized root mean square error (NRMSE) in estimating the reflection coefficient β_{x_2} for different methods as a function of the parameter k.

In Figs. 8 and 9 the x-axis regards to the power of Laplacian for the proposed method or number of nearest neighbor in KNN-search method. As it can be seen from these Figs. 8 and 9, the proposed method outperforms the KNN-search method for large enough k in estimating the parameters of acoustic impulse response. It is worth mentioning that proposed method can be easily utilized to recover the direction of arrival of a random source in a room based on observations from a *single* microphone and training.

4. Conclusions and discussions

We have proposed a novel technique for supervised learning when the data lies on a manifold. We treated this problem as a problem of extending a band-limited function on a homogeneous manifold while the function and the manifold are both unknown. The only assumption for reconstructing the function is that we have enough labeled data to reconstruct the function (i.e. enough data points on the manifold on with known function values) and enough data to approximate the Laplace–Beltrami operator on the manifold. Diffusion maps were utilized in order to learn the manifold (i.e. approximating the Laplace–Beltrami operator on the manifold). Simulation results show the advantage of the proposed method over the widely-used geometric harmonics. The proposed method is insensitive to the sampling procedure on the manifold. More specifically, the proposed method is not sensitive to the distribution of points sampled on the manifold. This is related to the property that diffusion maps with appropriate normalization recovered the Riemannian geometry of the data set.

An important issue which must be addressed here is the obtaining the constant c_0 (see condition (c) above). This constant is manifold-dependent and must be obtained appropriately to know if we have enough data for perfect reconstruction. The question of how to obtain this constant cannot be answered in general except, that for a compact manifold of dimension d the number of sampling points should be of order $\omega^{d/2}$, where ω is the bandwidth. See [5] for further discussion.

The last important issue which is interesting and needs more research is the behavior of the proposed method in the case of $L^{\infty}(\mathcal{M})$ norm. Answering to this question is beyond the scope of this paper and to the best of our knowledge it is an open problem. We conjecture that under this norm of the error is bounded. A special case of this problem has been answered in [33].

Acknowledgments

The authors thank the anonymous reviewers for their helpful comments. This research was supported by the Israel Science Foundation (Grant no. 1130/11).

References

[1] M. Mohri, A. Rostamizadeh, A. Talwalkar, Foundations of Machine Learning, The MIT Press, Cambridge, 2012.
[2] C.E. Shannon, Communication in the presence of noise, Proc. IRE 37 (1) (1949) 10–21.
[3] I. Schoenberg, Cardinal Spline Interpolation, CBMS, vol. 12, SIAM, Philadelphia, 1973.
[4] Y. Lyubarskii, W.R. Madych, The recovery of irregularly sampled band limited functions via tempered splines, J. Funct. Anal. 125 (1994) 201–222.
[5] I. Pesenson, A sampling theorem on homogeneous manifolds, Trans. Am. Math. Soc. 352 (9) (2000) 4257–4269.
[6] R.R. Coifman, S. Lafon, Geometric harmonics: a novel tool for multiscale out-of-sample extension of empirical functions, Appl. Comput. Harmon. Anal. 21 (2006) 31–52.
[7] A. Jansen, P. Niyogi, A Geometric Perspective on Speech Sounds, Tech. Rep., Computer Science Dept., Univ. of Chicago., 2005.
[8] D.L. Donoho, C. Grimes, Image manifolds which are isometric to euclidean space, J. Math. Imaging Vis. 23 (1) (2005) 5–24.
[9] C. Fowlkes, S. Belongie, F. Chung, J. Malik, Spectral grouping using the Nyström method, IEEE Trans. Pattern Anal. Mach. Intell. 26 (2004) 214–225.
[10] L. Zelnik-Manor, P. Perona, Self-tuning spectral clustering, Adv. Neural Inf. Process. Syst. 17 (2005) 1601–1608.
[11] S. Gepshtein, Y. Keller, Image completion by diffusion maps and spectral relaxation, IEEE Trans. Image Process. 22 (2013) 2983–2994.
[12] R. Talmon, I. Cohen, S. Gannot, Transient noise reduction using nonlocal diffusion filters, IEEE Trans. Audio Speech Lang. Process. 19 (2011) 1584–1599.
[13] S. Mousazadeh, I. Cohen, Voice activity detection in presence of transient noise using spectral clustering, IEEE Trans. Audio Speech Lang. Process. 21 (6) (2013) 1261–1271.
[14] A. Singer, Spectral independent component analysis, Appl. Comput. Harm. Anal. 21 (1) (2006) 135–144.
[15] A. Singer, R. Coifmain, Non-linear independent component analysis with diffusion maps, Appl. Comput. Harmon. Anal. 25 (2) (2008) 226–239.
[16] R. Talmon, D. Kushnir, R.R. Coifman, I. Cohen, S. Gannot, Parametrization of linear systems using diffusion kernels, IEEE Trans. Signal Process. 60 (3) (2012) 1159–1173.
[17] R. Talmon, I. Cohen, S. Gannot, Supervised source localization using diffusion kernels, in: Proceedings of 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA-2011, New Paltz, NY, 2011, pp. 16–19.
[18] X. Zhu, Z. Ghahramani, J. Lafferty, Semi-supervised learning using gaussian fields and harmonic functions, in: International Conference on Machine Learning, vol. 3, 2003, pp. 912–919.
[19] B. Nadler, N. Srebro, X. Zhou, Statistical analysis of semi-supervised learning: the limit of infinite unlabelled data, in: Advances in Neural Information Processing Systems, 2009, pp. 1330–1338.
[20] X. Zhou, M. Belkin, Semi-supervised learning by higher order regularization, in: International Conference on Artificial Intelligence and Statistics, 2011, pp. 892–900.

[21] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2319–2323.
[22] B. Scholkopf, A. Smola, K. Muller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (1996) 1299–1319.
[23] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (2003) 1373–1396.
[24] R.R. Coifman, S. Lafon, Diffusion maps, Appl. Comput. Harmon. Anal. 21 (2006) 5–30.
[25] B. Nadler, S. Lafon, R.R. Coifman, I.G. Kevrekidis, Diffusion maps, spectral clustering and reaction coordinates of dynamical systems, Appl. Comput. Harmon. Anal. 21 (1) (2006) 113–127.
[26] D.G. Luenberger, Introduction to linear and nonlinear programming, Vol. 28, Addison-Wesley, Reading, MA, 1973.
[27] M. Hein, J.-Y. Audibert, Intrinsic dimensionality estimation of submanifolds in R^d, in: ICML, 2005, pp. 289–296.
[28] R. Coifman, Y. Shkolnisky, F.J. Sigworth, A. Singer, Graph Laplacian tomography from unknown random projections, IEEE Trans. Image Process. 17 (2008) 1891–1899.
[29] M.I. Kadec, The exact value of the Paley-Wiener constant, Sov. Math. Dokl. 5 (1964) 559–561.
[30] P. Stoica, R.L. Moses, Introduction to Spectral Analysis, Prentice Hall, Upper Saddle River, NJ, 1997 Vol. (1).
[31] J.B. Allen, D.A. Berkley, Image method for efficiently simulating small-room acoustics, J. Acoust. Soc. Am. 65 (1979) 943.
[32] E.A.P. Habets, Room impulse response (RIR) generator, Sep. 2010. URL ⟨http://home.tiscali.nl/ehabets/rirgenerator.html⟩.
[33] A.Y. Shadrin, The L^∞-norm of the 2-spline projector is bounded independently of the knot sequence: a proof of de boor's conjecture, Acta Math. 187 (1) (2001) 59–137.