Applied Mathematics and Computation 000 (2017) 1–13

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

Multi-scale kernels for Nyström based extension schemes

N. Rabin*, D. Fishelov

Department of Mathematics, Afeka- Tel-Aviv Academic College of Engineering, Tel-Aviv, Israel

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Kernel methods
Manifold learning
Dimensionality reduction
Function extension

ABSTRACT

Nonlinear dimensionality reduction methods often include the construction of kernels for embedding the high-dimensional data points. Standard methods for extending the embedding coordinates (such as the Nyström method) also rely on spectral decomposition of kernels. It is desirable that these kernels capture most of the data sets' information using only a few leading modes of the spectrum.

In this work we propose multi-scale kernels, which are constructed as combinations of Gaussian kernels, to be used for kernel-based extension schemes. We review the kernels' spectral properties and show that their first few modes capture more information compared to the standard Gaussian kernel. Their application is demonstrated on a synthetic data-set and also applied to a real-life example that models daily electricity profiles and predicts the average day-ahead behavior.

1. Introduction

Nonlinear dimensionality reduction methods [3,4,7,26,34] often include constructions of kernels that describe pairwise distances between data points. The dimensionality reduction is achieved by a spectral decomposition of these kernels. The first few eigenvectors embed the data into a low-dimensional space. Extending the embedding coordinates to new data points is typically carried out by kernel-based techniques, such as the Nyström method [1,5]. For the efficiency of these methods, it is desirable that the kernels used for the extension capture most of the datasets' information using a few leading modes of the spectrum.

In this work we propose multi-scale kernels, which are constructed as combinations of Gaussian kernels, to be used within a kernel-based extension framework. The multi-scale kernels are high-order kernels that are often used as convolution kernels for function approximations. The proposed kernels were suggested, for example, in the context of fluid-dynamics simulations [2,15]. In the context of image processing, these kernels are related to forward-backward diffusion processes [16,29], which preserve edge information while smoothing the data. In this paper we show that the low-order spectral modes of the multi-scale kernels capture more information compared to the standard Gaussian kernel. This results in a more accurate procedure for extension of an embedded data set. In addition, these kernels may be represented as polynomials of matrices related to a Gaussian kernel, where the construction is independent of the dimension of the data. This property facilitates the applications of these kernels in any dimension.

This paper is organized as follows. Section 2 describes a framework for modeling high-dimensional datasets by a low-dimensional manifold. In addition this section describes methods for extending the low-dimensional representation to new

* Corresponding author.
E-mail addresses: netar@afeka.ac.il, neta.rabin@gmail.com (N. Rabin).

http://dx.doi.org/10.1016/j.amc.2017.02.025
0096-3003/

Algorithm 1 Diffusion Maps.

Input: $\Gamma = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\} \subseteq \mathbb{R}^n$, the input dataset. Parameter: δ.

Output: A set of embedding coordinates $\{\Psi(\mathbf{x}_i)\}_{i=1}^{M}$.

1: Construct a graph $G = (\Gamma, W)$ with a kernel

$$W_{i,j} = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\delta^2}}.$$

2: Construct a normalized kernel

$$A_{i,j} = \frac{W_{i,j}}{(\sum_j W_{i,j}) \cdot (\sum_i W_{i,j})}.$$

3: Compute the normalized graph Laplacian

$$K_{i,j} = \frac{A_{i,j}}{\sum_j A_{i,j}}.$$

4: Compute the spectral decomposition of K:

$$K_{i,j} = \sum_{l \geq 0} \lambda_l \psi_l(\mathbf{x}_i) \phi_l(\mathbf{x}_j), \tag{1}$$

where $\{\lambda_l\}_{l=0}^{M-1}$ are the eigenvalues of K and $\{\phi_l\}_{l=0}^{M-1}$ and $\{\psi_l\}_{l=0}^{M-1}$ are the corresponding left and right eigenvectors of K.

5: Define the diffusion maps by

$$\Psi(\mathbf{x}_i) = (\lambda_1 \psi_1(\mathbf{x}_i), \lambda_2 \psi_2(\mathbf{x}_i), \lambda_3 \psi_3(\mathbf{x}_i) \cdots).$$

Algorithm 1. Diffusion Maps.

data points. Section 3 describes the construction of multi-scale kernels in $\mathbb{R}^2$ and in $\mathbb{R}^n$. Section 4 reviews the spectral properties of these kernels. In Section 5, we construct dimensional independent matrix representations of the newly-constructed kernels. Section 6 demonstrates the application of the proposed kernels within a Nyström based extension method.

2. Diffusion maps and geometric harmonics

Modeling high-dimensional datasets with dimensionality reduction methods has become a common practice in many fields such as signal processing [18,28,31], medical applications [12,14], genome modeling [30], climate modeling [6] and process control [23,33], to name some. The constructed low-dimensional description enable us to gain insight into our understanding of the phenomenon that generates and governs the inspected data. Analysis is then carried out in the new low-dimensional space.

Classical dimensionality reduction techniques, such as Principal Component Analysis [32] or Multidimensional Scaling [9], fail on datasets that have a complex geometric structure with a low intrinsic dimensionality. Recently, a great deal of attention has been directed to the so-called kernel methods, such as Local Linear Embedding [26], Laplacian Eigenmaps [3,4], Hessian Eigenmaps [11] and Local Tangent Space Alignment [34]. These algorithms exhibit two major advantages over classical methods: they are nonlinear and they preserve locality.

In the diffusion maps framework [7], which is used in this work, the ordinary Euclidean distance in the embedding space measures intrinsic diffusion distances in the original high-dimensional space. Geometric harmonics [8], which is a kernel based extension scheme, is then used to extend the model to include new data points. Sections 2.1 and 2.2 review the diffusion maps and geometric harmonics methods.

2.1. Diffusion maps

In diffusion maps one builds a graph of the data using symmetric and positive definite kernels. Algorithm 1 summarizes the steps of constructing diffusion maps, based on the spectral decomposition of the graph. The scale of the kernel δ is a parameter and may be set as described in [25].

As was shown in [7], the diffusion maps embed the data into a Euclidean space by preserving a pairwise distance $D_{i,j}$, where

$$D_{i,j}^2 = \sum_m \frac{(K_{i,m} - K_{m,j})^2}{\phi_0(\mathbf{x}_m)},$$

where $\phi_0(\mathbf{x}_m)$ is the first left eigenvector of K, it is also the stationary distribution of the Markov matrix K. This distance is called the *diffusion distance*. Substituting the spectral decomposition of K ([Eq.](1), Step 4 in Algorithm 1) in the definition

Please cite this article as: N. Rabin, D. Fishelov, Multi-scale kernels for Nyström based extension schemes, Applied Mathematics and Computation (2017), http://dx.doi.org/10.1016/j.amc.2017.02.025

Algorithm 2 Geometric Harmonics.

Input: $\Gamma = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\} \subseteq \mathbb{R}^n$, the input dataset, and $\Gamma \subset \bar{\Gamma}$. A function $\mathbf{f}$ given on Γ. Parameters: σ, η.

Output: Extension of f to $\bar{\Gamma}$.

1: Construct a kernel matrix (see Equation (26)), where σ is the extension scale.

2: Compute the orthonormal spectral decomposition of the kernel

$$\mu_l \varphi_l(\mathbf{x}_i) = \sum_{j=1}^{M} k_\sigma(\mathbf{x}_i - \mathbf{x}_j)\varphi_l(\mathbf{x}_j), \quad where \quad \mathbf{x}_i \in \Gamma.$$

3: Approximate f on Γ by

$$f(\mathbf{x}_i) = \sum_{l:\mu_l \geq \eta\mu_0} \langle \varphi_l, \mathbf{f} \rangle \, \varphi_l(\mathbf{x}_i), \quad where \quad \mathbf{x}_i \in \Gamma.$$

4: Extend φ_l to a new point $\mathbf{y} \in \bar{\Gamma}$ by

$$\bar{\varphi}_l(\mathbf{y}) = \frac{1}{\mu_l} \sum_{j=1}^{M} k_\sigma(\mathbf{y} - \mathbf{x}_j)\varphi_l(\mathbf{x}_j), .$$

5: Extend f to the new point $y \in \bar{\Gamma}$

$$\bar{f}(\mathbf{y}) = \sum_{l:\mu_l \geq \eta\mu_0} \langle \varphi_l, \mathbf{f} \rangle \, \bar{\varphi}_l(\mathbf{y}).$$

Algorithm 2. Geometric Harmonics.

of $D^2_{i,j}$ yields

$$D^2_{i,j} = \sum_l \lambda_l^2 \left(\psi_l(\mathbf{x}_i) - \psi_l(\mathbf{x}_j)\right)^2.$$

Thus, diffusion maps preserve pairwise distances.

2.2. Geometric harmonics

Geometric harmonics [8] were introduced along with diffusion maps as a complementary method for data extension. It involves the construction of harmonic functions. These functions form a basis, which may span a general function and in particular they are used to expand the diffusion maps coordinates. The construction of the geometric harmonics includes spectral decomposition of positive semi-definite kernel. Since the spectrum of such kernels decays fast to zero, only the leading terms, which are associated with large enough eigenvalues, are used in the scheme. Thus, we are interested in kernels that capture most of the information in the first few modes.

In [8] several kernels are discussed. The prolate spheroidal wave functions

$$K_B^{(n)}(r) = (B/2)^{n/2} \frac{J_{n/2}(\pi B r)}{r^{n/2}},$$

suggested in [22], are tested for extending the functions $f = \cos(j\theta)$, $j = 1, 2, 4, 8$ from the unit circle to $\mathbb{R}^2$. It was found that they tend to generate a large amount of oscillations outside the unit circle. In addition, this kernel needs the evaluation of a Bessel function, and it therefore requires substantial computational effort. Harmonic extension kernels, defined by

$$k(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} -\log\left(\|\mathbf{x}_i - \mathbf{x}_j\|\right) & n = 2 \\ \frac{1}{\|\mathbf{x}_i - \mathbf{x}_j\|^{n-2}}, & n \geq 3 \end{cases}$$

are also reviewed in [8]. These kernels do not involve a scale parameter, thus they are not localized in space. In addition, for oscillatory functions, their extension decays fast, limiting the width of the extension domain. Gaussian extensions were also tested in [8] . They are easy to compute and they are much more localized compared with the prolate spheroidal wave functions. The results shown in [8] demonstrate that the extension band for functions with high frequencies is wider compared with the extension width of harmonic extension kernels.

Algorithm 2 reviews the procedure of extending a function using geometric harmonics. Two parameters σ and η should be defined. The parameter σ is the extension scale of the kernel and it is chosen to be larger than δ used in Diffusion Maps Algorithm 1 (see [8,19]). The normalization of the Diffusion Maps kernel may consist of two steps (see Steps 2 and 3 in Algorithm 1), the kernel used in Geometric Harmonics has no normalization or just a row-sum normalization. Since the Diffusion Maps kernel and the Geometric Harmonics kernels are not identical, we will denote their spectral decomposition

Please cite this article as: N. Rabin, D. Fishelov, Multi-scale kernels for Nyström based extension schemes, Applied Mathematics and Computation (2017), http://dx.doi.org/10.1016/j.amc.2017.02.025

components with different letters: $\{\mu_l\}_{l=1}^M$ for the eigenvalues and $\{\phi_l\}_{l=1}^M$ for the eigenvectors. The eigenvalues of K, $\{\mu_l\}_{l=1}^M$, tend to zero as $l \to \infty$. Therefore, by using the eigenvectors satisfying the relation $\mu_l \geq \eta\mu_0$, one avoids an ill-conditioned procedure. This algorithm is used for extending the diffusion maps embedding coordinates, i.e., the input functions $\mathbf{f}$ are the embedding coordinates that are constructed in Step 5 of Algorithm 1.

3. Construction of multi-scale kernels

This section describes the construction of the multi-scale kernels using combinations of scaled Gaussians. When used as convolution kernels for numerical analysis applications, these kernels achieve high order approximations. High-order kernels are constructed by requiring several moment conditions on the kernels. The more moment conditions are satisfied the smaller is the difference between f and $k_\delta * f$. We show in this section that the difference $\|f * k_{2,\delta} - f\|_{L^2}$ is bounded by $C\|f^{(2)}\|_{L^2}\delta^2$. Thus, if more moment conditions are satisfied then k_δ is closer to a delta function. When compared to low-order kernels, the spectrum of such a kernel preserves more energy in the leading eigenvalues. In the context of Nyström based extensions, this property is essential, since only the leading eigenvalues are involved in the extension process.

Let $X \subset \mathbb{R}^n$ be a set of data points. We define a kernel $k : X \to \mathbb{R}$, that satisfies the following moment conditions.

$$\begin{aligned} &\text{(i)} \quad \int k(\mathbf{x})d\mathbf{x} = 1, \\ &\text{(ii)} \quad \int \mathbf{x}^\beta k(\mathbf{x})d\mathbf{x} = 0, \qquad 1 \leq |\beta| \leq d-1. \\ &\text{(iii)} \quad \int |\mathbf{x}^\beta k(\mathbf{x})|d\mathbf{x} \leq C, \qquad |\beta| = d. \end{aligned} \tag{2}$$

We confine ourselves to kernels which depend on $r = |\mathbf{x}|$, thus the kernels are symmetric. In this case the requirements (ii) of Eq. (2) hold for $|\beta|$ odd, thus assuming that d is even. A natural choice of a kernel in $\mathbb{R}^n$ is

$$k(\mathbf{x}) = \frac{1}{\pi^{n/2}} e^{-r^2}, \tag{3}$$

where the constant $1/\pi^{n/2}$ is chosen such that the requirement $\int k(\mathbf{x})d\mathbf{x} = 1$ is met. Since this kernel is radially symmetric, we have $\int \mathbf{x}\, k(\mathbf{x})d\mathbf{x} = 0$ and thus Eq. (2) is satisfied with $d = 2$.

Define k_δ by

$$k_\delta(\mathbf{x}) = \frac{1}{\delta^n} k\left(\frac{\mathbf{x}}{\delta}\right).$$

Using the kernel k_δ, one may approximate a function $f : X \to \mathbb{R}$ by

$$(k_\delta * f)(\mathbf{x}) = \int k_\delta(\mathbf{x} - \mathbf{x}')f(\mathbf{x}')\, d\mathbf{x}'. \tag{4}$$

The integral on the right-hand side is approximated by the sum $\sum_j k_\delta(\mathbf{x} - \mathbf{x}_j)f(\mathbf{x}_j)\, \Delta\mathbf{x}_j$, where $\Delta\mathbf{x}_j$ is a volume element in R^n. Choosing the kernel k such that more moment conditions are satisfied yields a higher order approximation for f.

3.1. Multi-scale kernels in $\mathbb{R}^2$

For simplicity, we start by constructing kernels in $\mathbb{R}^2$. We construct kernels k_2, k_4, k_6 of order 2, 4 and 6 respectively. First, we obtain a second-order kernel by using a Gaussian kernel and then show that specific linear combinations of scaled Gaussians result in higher-order kernels.

3.1.1. A second-order Gaussian kernel in $\mathbb{R}^2$

Choosing $n = 2$ in Eq. (3) yields $k_2(r) = \frac{1}{\pi} e^{-r^2}$. One may readily check that k_2 satisfies

$$\begin{aligned} &\int k_2(\|x\|)d\mathbf{x} = 1, \\ &\int \mathbf{x}k_2(\|x\|)d\mathbf{x} = 0, \\ &\int |\mathbf{x}^2 k_2(\|x\|)|d\mathbf{x} \leq C. \end{aligned} \tag{5}$$

In Appendix A we prove that is kernel is a second order kernel. Thus, given the kernel $k_{2,\delta}(r) = \frac{1}{\delta^2}k_2(r/\delta)$ and a function $f : X \to \mathbb{R}$, $(X \subset \mathbb{R}^2)$, we have

$$\|f * k_{2,\delta} - f\|_{L^2} \leq C\|f^{(2)}\|_{L^2}\delta^2.$$

The above inequality is correct for any function f satisfying $f^{(2)} \in L_2$, where $f^{(2)}$ denotes all second-order derivatives of f.

Please cite this article as: N. Rabin, D. Fishelov, Multi-scale kernels for Nyström based extension schemes, Applied Mathematics and Computation (2017), http://dx.doi.org/10.1016/j.amc.2017.02.025

3.1.2. A fourth-order kernel in $\mathbb{R}^2$

Now we proceed to the construction of a fourth-order kernel in $\mathbb{R}^2$. Let

$$k_4(r) = Ak_2(r) + Bk_2(r/a). \tag{6}$$

Require that $\int k_4(r)d\mathbf{x} = 1$, thus $A + a^2B = 1$. For the kernel k_4 to satisfy $\int \mathbf{x}^2 k_4(r)d\mathbf{x} = 0$, we need to have $A + a^4B = 0$. Choosing $a^2 = 2$, results in the following set of requirements

$$\begin{aligned} A + 2B &= 1, \\ A + 4B &= 0. \end{aligned} \tag{7}$$

This yields $A = 2, \quad B = -1/2$. Therefore, a fourth-order kernel in $\mathbb{R}^2$ is (see [15])

$$k_4(r) = \frac{1}{2\pi}(4\ e^{-r^2} - e^{-r^2/2}).$$

Define $k_{4,\delta}(r) = \frac{1}{\delta^2}k_4(r/\delta)$, then if $f^{(4)}$ belongs to L^2 we have

$$\|f * k_{4,\delta} - f\|_{L^2} \le C\|f^{(4)}\|_{L^2}\delta^4.$$

3.1.3. A sixth-order kernel in $\mathbb{R}^2$

In this subsection we construct a sixth-order kernel in $\mathbb{R}^2$. Let

$$k_6(r) = Ak_2(r) + Bk_2(r/a) + Ck_2(r/a^2). \tag{8}$$

Requiring $\int k_6(r)d\mathbf{x} = 1$, yields $A + a^2B + (a^2)^2C = 1$. In addition, in order that $\int \mathbf{x}^2 k_6(r)d\mathbf{x} = 0$, one needs to assure that $A + a^4B + (a^2)^4C = 0$. Requiring that $\int \mathbf{x}^4 k_6(r)d\mathbf{x} = 0$, results in $A + a^6B + (a^2)^6C = 0$. Choosing $a^2 = 2$ results in the following set of requirements.

$$\begin{aligned} A + 2B + 4C &= 1, \\ A + 4B + 16C &= 0, \\ A + 8B + 64C &= 0. \end{aligned} \tag{9}$$

This yields $A = 32/12, \quad B = -1, \quad C = 1/12$. Therefore,

$$k_6(r) = \frac{1}{12\pi}(32\ e^{-r^2} - 12\ e^{-r^2/2} + e^{-r^2/4}).$$

Define $k_{6,\delta}(r) = \frac{1}{\delta^2}k_6(r/\delta)$. The error related to this kernel is bounded by

$$\|f * k_{6,\delta} - f\|_{L^2} \le C\|f^{(6)}\|_{L^2}\delta^6.$$

3.2. High-order kernels in $\mathbb{R}^n$

We now construct high-order kernels in $\mathbb{R}^n$. In this case the kernels are also combinations of scaled Gaussians and their coefficients depend on the dimension n.

3.2.1. A second-order Gaussian kernel in $\mathbb{R}^n$

A second order kernel in $\mathbb{R}^n$ is (see Eq. (3))

$$k_2(r) = \frac{1}{\pi^{n/2}}e^{-r^2}. \tag{10}$$

Defining $k_{2,\delta}(r) = \frac{1}{\delta^n}k_2(r/\delta)$ yields

$$\|f * k_{2,\delta} - f\|_{L^2} \le C\|f^{(2)}\|_{L^2}\delta^2.$$

3.2.2. A fourth-order kernel in $\mathbb{R}^n$

A fourth-order kernel in $\mathbb{R}^n$ is constructed using two Gaussians

$$k_4(r) = Ak_2(r) + Bk_2(r/a).$$

We require that $\int k_4(r)d\mathbf{x} = 1$, and $\int \mathbf{x}^2 k_4(r)d\mathbf{x} = 0$, thus $A + a^nB = 1$ and $A + a^{n+2}B = 0$. Choosing $a^2 = 2$, we have the following set of requirements

$$\begin{aligned} A + (\sqrt{2})^nB &= 1, \\ A + (\sqrt{2})^{n+2}B &= 0. \end{aligned} \tag{11}$$

This yields $A = \frac{2\cdot 2^{n/2}}{2^{n/2}}, \quad B = -\frac{1}{2^{n/2}}$. Therefore, a fourth-order kernel in $\mathbb{R}^n$ is

$$k_4(r) = \frac{1}{(2\pi)^{n/2}}(2\cdot 2^{n/2}\ e^{-r^2} - e^{-r^2/2}).$$

If $f^{(4)}$ is in L^2 then $\|f * k_{4,\delta} - f\|_{L^2} \le C\|f^{(4)}\|_{L^2}\delta^4$, where $k_{4,\delta} = \frac{1}{\delta^n}k_4(r/\delta)$.

Please cite this article as: N. Rabin, D. Fishelov, Multi-scale kernels for Nyström based extension schemes, Applied Mathematics and Computation (2017), http://dx.doi.org/10.1016/j.amc.2017.02.025

3.2.3. A sixth-order kernel in $\mathbb{R}^n$

Combinations of three scaled Gaussians are used to construct a six-order kernel. Thus,

$$k_6(r) = Ak_2(r) + Bk_2(r/a) + Ck_2(r/a^2).$$

Requiring $\int k_6(r)d\mathbf{x} = 1$, $\int \mathbf{x}^2 k_6(r)d\mathbf{x} = 0$, and $\int \mathbf{x}^4 k_6(r)d\mathbf{x} = 0$, results in

$$\begin{aligned} &A + a^n B + (a^2)^n C = 1, \\ &A + a^{n+2} B + (a^2)^{n+2} C = 0, \\ &A + a^{n+4} B + (a^2)^{n+4} C = 0. \end{aligned} \tag{12}$$

Choosing $a^2 = 2$ yields

$$\begin{aligned} &A + (\sqrt{2})^n B + 2^n C = 1, \\ &A + (\sqrt{2})^{n+2} B + 2^{n+2} C = 0, \\ &A + (\sqrt{2})^{n+4} B + 2^{n+4} C = 0. \end{aligned} \tag{13}$$

Thus, $A = 8 \cdot 2^n/(3\ 2^n)$, $B = -6\ (\sqrt{2})^n/(3 \cdot 2^n)$, $C = 1/(3\ 2^n)$. Therefore,

$$k_6(r) = \frac{1}{\pi^{n/2}} \frac{1}{3 \cdot 2^n} (8 \cdot 2^n\ e^{-r^2} - 6\ (\sqrt{2})^n\ e^{-r^2/2} + e^{-r^2/4}).$$

The error obtained by this kernel is bounded by

$$\|f * k_{6,\delta} - f\|_{L^2} \le C \|f^{(6)}\|_{L^2} \delta^6,$$

where $k_{6,\delta} = \frac{1}{\delta^n} k_6(r/\delta)$.

4. Spectral properties of high order kernels

An important property of kernel based methods is the decay rate of the spectrum associated with the kernels. It is desirable that only a few eigenfunctions capture most of the information. This means that when starting with the largest eigenvalue 1, only a few eigenvalues are close to 1, while the rest of the eigenvalues decay fast to zero.

Let $k : X \to \mathbb{R}$ be a convolution kernel. Its eigenvalue λ and the corresponding eigenfunction ϕ satisfy

$$k * \phi = \lambda\phi. \tag{14}$$

Taking the Fourier transform of Eq. (14), we have $\hat{k}(\omega)\hat{\phi}(\omega) = \lambda\hat{\phi}(\omega)$, thus

$$\hat{k}(\omega) = \lambda. \tag{15}$$

Therefore the spectrum of k is $\hat{k}(\omega)$.

Lemma 4.1 (The spectrum of $k_2(r)$)**.** *Consider the second order kernel in $\mathbb{R}^n$*

$$k_2(r) = \frac{1}{\pi^{n/2}} e^{-r^2}.$$

Let its Fourier transform be

$$\hat{k_2}(\omega) = \frac{1}{\pi^{n/2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\sum_{j=1}^n x_j^2} e^{-i\sum_{j=1}^n \omega_j x_j} dx_1 \cdots dx_n, \tag{16}$$

where $\omega = (\omega_1, \ldots, \omega_n)$. Then

$$\hat{k_2}(s) = e^{-s^2/4} > 0, \tag{17}$$

where $s^2 = \omega_1^2 + \omega_2^2 \cdots + \omega_n^2$.

Proof. The definition (16) of $\hat{k_2}(\omega)$, together with the identity

$$\int_{-\infty}^{\infty} e^{-x_1^2} e^{-i\omega_1 x_1} dx_1 = \sqrt{\pi} e^{-\omega_1^2/4}, \tag{18}$$

yield

$$\hat{k_2}(\omega) = e^{-(\omega_1^2 + \omega_2^2 \cdots + \omega_n^2)/4}.$$

Therefore (with a slight abuse of notations),

$$\hat{k_2}(s) = e^{-s^2/4},$$

where $s^2 = \omega_1^2 + \omega_2^2 \cdots + \omega_n^2$. □

Note that $\hat{k}_2(s = 0) = 1$ and that $\hat{k}_2(s)$ decreases from 1 to zero (for $s \ge 0$) as s tends to ∞.

Please cite this article as: N. Rabin, D. Fishelov, Multi-scale kernels for Nyström based extension schemes, Applied Mathematics and Computation (2017), http://dx.doi.org/10.1016/j.amc.2017.02.025

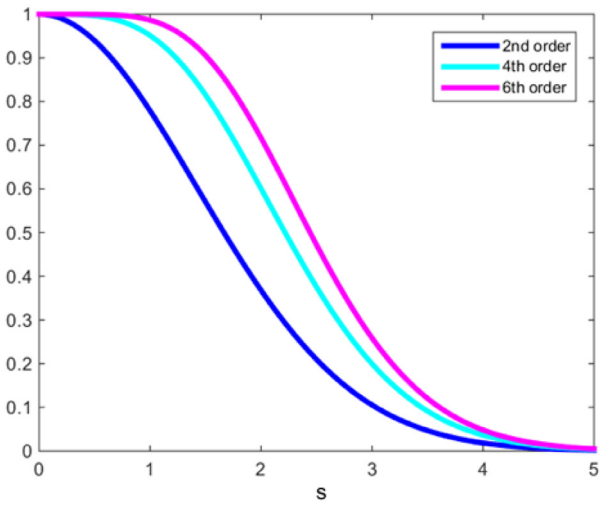

Fig. A.1. Spectral decay for second-order, fourth-order and sixth-order kernels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Lemma 4.2 (The spectrum of $k_4(r)$)**.** *The spectrum of the fourth-order kernel in* $\mathbb{R}^n$

$$k_4(r) = \frac{1}{(2\pi)^{n/2}}[2\cdot 2^{n/2}e^{-r^2} - e^{-r^2/2}]$$

is

$$\hat{k_4}(s) = 2e^{-s^2/4} - e^{-s^2/2} = e^{-s^2/4}(2 - e^{-s^2/4}). \tag{19}$$

In addition it satisfies

$$\hat{k_4}(s) \geq \hat{k_2}(s) > 0. \tag{20}$$

Proof. The proof of (19) is carried out using the definition of $\hat{k}_4$, a change of variable $y_j = x_j/\sqrt{2}$ in the second term of the integral related to $\hat{k}_4$, and the identity (18). In addition, since $1 \leq 2 - e^{-s^2/4} < 2$, using (17) and (19), it follows that $\hat{k_4}(s) \geq \hat{k_2}(s) > 0$. □

Lemma 4.3 (The spectrum of $k_6(r)$)**.**

The spectrum of the sixth-order kernel in $\mathbb{R}^n$,

$$k_6(r) = \frac{1}{\pi^{n/2}}\frac{1}{3\cdot 2^n}(8\cdot 2^n\ e^{-r^2} - 6\ (\sqrt{2})^n\ e^{-r^2/2} + e^{-r^2/4})$$

is

$$\hat{k_6}(s) = \frac{1}{12}[32e^{-s^2/4} - 24e^{-s^2/2} + 4e^{-s^2}] = \frac{1}{3}e^{-s^2/4}[8 - -6e^{-s^2/4} + e^{-3s^2/4}]. \tag{21}$$

In addition

$$\hat{k_6}(s) \geq \hat{k_4}(s) \geq \hat{k_2}(s) > 0. \tag{22}$$

Proof. Eq. (21) is proved using the definition of $\hat{k}_6$, change of variables in the integral related to $\hat{k}_6$, and the identity (18). Since $8 - 6e^{-s^2/4} + e^{-3s^2/4} \geq 3(2 - e^{-s^2/4})$ and using the inequality (20), we deduce that

$$\hat{k_6}(s) \geq \hat{k_4}(s) \geq \hat{k_2}(s) > 0.$$

□

Notice that the Fourier transforms of the kernels do not depend on the dimension n. Moreover, the spectrum of k_2 decreases from one to zero faster than the spectrum of k_4 and the spectrum of k_4 decreases from one to zero faster than the spectrum of k_6. Thus, the eigenvalues of k_6 in some range $0 \leq s \leq S_0$ capture more information that those of k_4 in same range. In addition, the eigenvalues of k_4 in some range $0 \leq s \leq S_0$ capture more information that those of k_2 in same range. This is illustrated in Fig. A.1.

Repeating the calculations above for $k_{2,\delta}$, $k_{4,\delta}$ and $k_{6,\delta}$, yields the following formulas for the spectrum.

$$\begin{aligned}
&\hat{k}_{2,\delta}(s) = e^{-s^2\delta^2/4},\\
&\hat{k}_{4,\delta}(s) = 2e^{-s^2\delta^2/4} - e^{-s^2\delta^2/2}\\
&\hat{k}_{6,\delta}(s) = \tfrac{1}{3}(8e^{-s^2\delta^2/4} - 6e^{-s^2\delta^2/2} + e^{-s^2\delta^2}).
\end{aligned} \tag{23}$$

Please cite this article as: N. Rabin, D. Fishelov, Multi-scale kernels for Nyström based extension schemes, Applied Mathematics and Computation (2017), http://dx.doi.org/10.1016/j.amc.2017.02.025

In addition, the following inequality holds.

$$\hat{k}_{6,\delta}(s) \geq \hat{k}_{4,\delta}(s) \geq \hat{k}_{2,\delta}(s) > 0. \tag{24}$$

We comment here that additional high-order kernels were suggested in the context of fluid dynamics. An infinite-order kernel in 2*D* was suggested by Hald [17]. It is given by

$$k(r) = \frac{4}{45\pi r^3}(16J_3(4r) - 10J_3(2r) + J_3(r)).$$

The Fourier transform of this kernel is given by

$$\hat{k}(s) = \begin{cases} 1, & 0 \leq s \leq 1 \\ 44 + 2s^2 - s^4, & 1 \leq s \leq 2 \\ 256 - 32s^2 + s^4, & 2 \leq s \leq 4 \\ 0, & s > 4 \end{cases}$$

This kernel needs three evaluation of a Bessel function, and it therefore requires more computational effort compared to the Gaussian kernels used in this paper. Nordmark [21] proposed the following eight-order cutoff function with compact support in 2*D*

$$k(r) = \begin{cases} \frac{52}{\pi}(140r^6 - 105r^4 + 21r^2 - 1)(1 - r^2)^9, & 0 \leq r \leq 1 \\ 0, & r > 1 \end{cases}$$

Its Fourier transform is

$$\hat{k}(s) = -6656 * 11!\pi^{-2}(s^{-10}J_{10}(s) - 84s^{-11}J_{11}(s) + 2520s^{-12}J_{12}(s) - 26880s^{-13}J_{13}(s)).$$

This kernel does not require expensive evaluations of functions, since it is constructed via polynomials. However, an eight-order kernel may be too fine to handle real-world data, thus for the examples in Section 6, it is preferable to use kernels up to sixth order.

5. Representation of the high-order kernels in matrix forms

For practical applications the integrals in the convolution (4) is replaced by a discrete sum. This means that $k_\delta * f$ is approximated by the sum

$$f(\mathbf{x}_j) \simeq \sum_{l=1}^{M} k_\delta(\mathbf{x}_j - \mathbf{x}_l) f(\mathbf{x}_l) \Delta\mathbf{x}_l, \quad j = 1, \ldots, M \tag{25}$$

where $\Delta\mathbf{x}_l$ is a volume of element *l*. Let $\mathbf{f} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_M))$ be the vector which contains the discrete values of *f* at the points $\mathbf{x}_1, \ldots, \mathbf{x}_M$. The right hand-side of (25) may be represented by $K\mathbf{f}$, where *K* is the matrix

$$K_{j,l} = k_\delta(\mathbf{x}_j - \mathbf{x}_l). \tag{26}$$

We show that in practice the matrix *K* does not depend on the dimension of the data. This yields an efficient construction of high-order kernels using polynomials of the second-order (normalized) Gaussian kernel. We construct polynomial representations for the fourth and the sixth order kernels.

Assume that $K_2\mathbf{f}$ is the matrix representation of $k_{2,\delta} * f$. Therefore, $K_2^2\mathbf{f}$ represents the convolution $k_2 * k_2 * f$. The Fourier transform of $k_2 * k_2 * f$ is $\hat{k_2} \cdot \hat{k_2} \cdot \hat{f} = (\hat{k_2})^2\hat{f}$. Thus, the application of the matrix K_2 twice on $\mathbf{f}$, i.e., $K_2^2\mathbf{f}$ corresponds to $(\hat{k_2})^2\hat{f}$. Using Eq. (17) we deduce that $(\hat{k_2})^2 = e^{-s^2/2}$. Thus, Eq. (19) can be written as

$$\hat{k_4} = 2e^{-s^2/4} - e^{-s^2/2} = 2\hat{k_2} - (\hat{k_2})^2.$$

We conclude that the matrix representation of k_4 is

$$K_4 = 2K_2 - K_2^2. \tag{27}$$

In a similar way, Eq. (21) can be written as

$$\hat{k_6} = \frac{1}{12}[32e^{-s^2/4} - 24e^{-s^2/2} + 4e^{-s^2}] = \frac{1}{3}[8\hat{k_2} - 6(\hat{k_2})^2 + (\hat{k_2})^4]. \tag{28}$$

The matrix representation of k_6 is therefore

$$K_6 = (8K_2 - 6K_2^2 + K_2^4)/3. \tag{29}$$

In the next section we combine the newly constructed kernels with a dimensionality-reduction framework.

Please cite this article as: N. Rabin, D. Fishelov, Multi-scale kernels for Nyström based extension schemes, Applied Mathematics and Computation (2017), http://dx.doi.org/10.1016/j.amc.2017.02.025

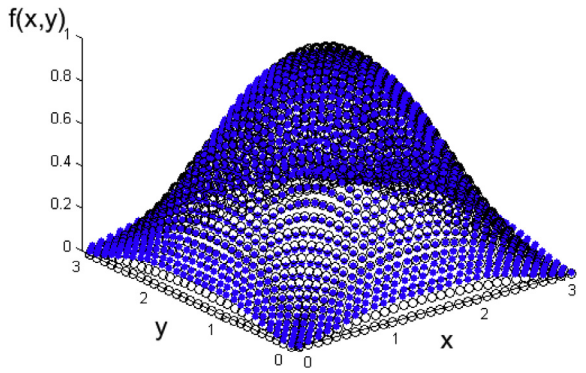

Fig. A.2. Approximation of $f(x, y) = \sin(x)sin(y),\ 0 \le x, y \le \pi$ using $k_{2,\delta}$, blue points are the approximations of the exact values in black. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

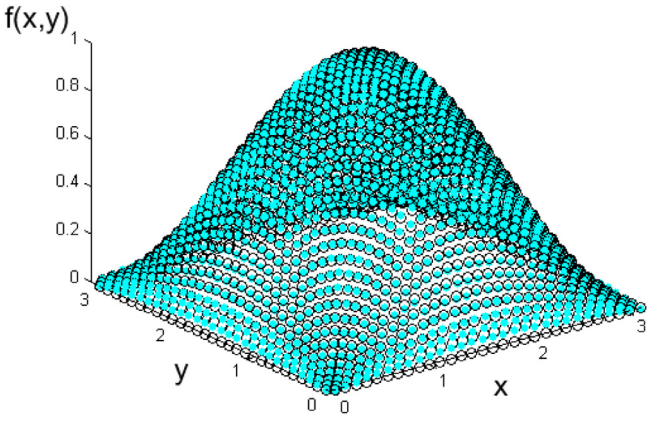

Fig. A.3. Approximation of $f(x, y) = \sin(x)sin(y),\ 0 \le x, y \le \pi$ using $k_{4,\delta}$, light blue points are the approximations of the exact values in black. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Experimental results

The following examples demonstrate the advantages of integrating high-order kernels in the Geometric Harmonics extension method. The scale δ in these experiments was computed by the following method that was proposed in [27]. Given a set of n data points, the pairwise Euclidean distance matrix between data points is defined as $D = \{dij\}_{i;\,j=1\ldots n}$. We set δ to be

$$\delta = 2 \times \text{median}\{d_{i,j}\}_{i;\,j=1\ldots n}.$$

This construction provides an estimate to the average pairwise distance.

6.1. Example 1

As an elementary example, consider the function $f(x, y) = \sin(x)\sin(y),\ 0 \le x, y \le \pi$. A uniform grid of $N \times N$ points (x_i, y_j) is laid out, where $x_i = (i-1)\pi/N$, and $y_j = (j-1)\pi/N$, $i, j = 1, \ldots, N+1$. Assume that the function is known on the points (x_i, y_j) for which $i + j$ is even. In order to evaluate the function on the rest of the points ($i + j$ odd), Algorithm 2 is applied with kernels of various orders. For this example N is set to 60 and $\delta = 0.3$.

Fig. A.2 shows the results for the approximation of $f(x, y) = \sin(x)\sin(y)$ using $k_{2,\delta}$. The discrete L_2 error is 3.187×10^{-2}, and the discrete L_∞ error is 1.071×10^{-1}. One may notice that the largest errors are obtained for near-boundary points.

Next, f is approximated at the set of new points using $k_{4,\delta}$. The results are given in Fig. A.3. The discrete L_2 error is 1.168×10^{-2}, and the discrete L_∞ error is 4.664×10^{-2}. It is seen that there is an improvement of the approximation at near-boundary points.

Last, the calculation of $f(x, y) = \sin(x)\sin(y)$ is repeated at the set of new points using $k_{6,\delta}$. The results are given in Fig. A.4. The discrete L_2 error is 7.48×10^{-3}, and the discrete L_∞ error is 3.571×10^{-2}. There is further improvement of the approximation at near-boundary points.

Fig. A.5 plots the spectrum of the three kernels (the spectrum is computed in Step 2 of Algorithm 2).

This behavior is in agreement with the theoretical computations of the spectrum (see Fig. A.1).

Please cite this article as: N. Rabin, D. Fishelov, Multi-scale kernels for Nyström based extension schemes, Applied Mathematics and Computation (2017), http://dx.doi.org/10.1016/j.amc.2017.02.025

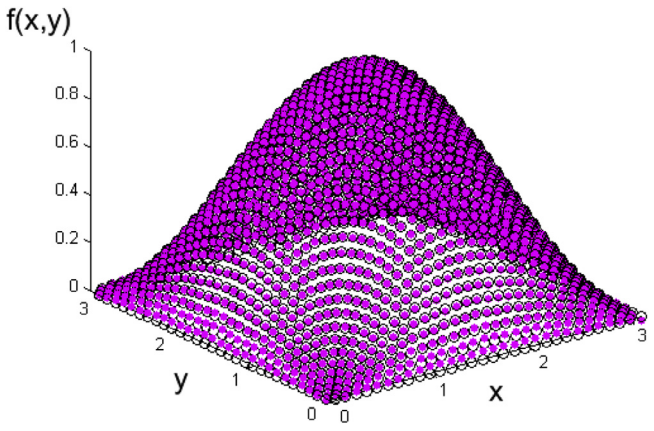

Fig. A.4. Approximation of $f(x, y) = \sin(x)sin(y)$, $0 \le x, y \le \pi$ using $k_{6,\delta}$, pink points are the approximations of the exact values in black. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

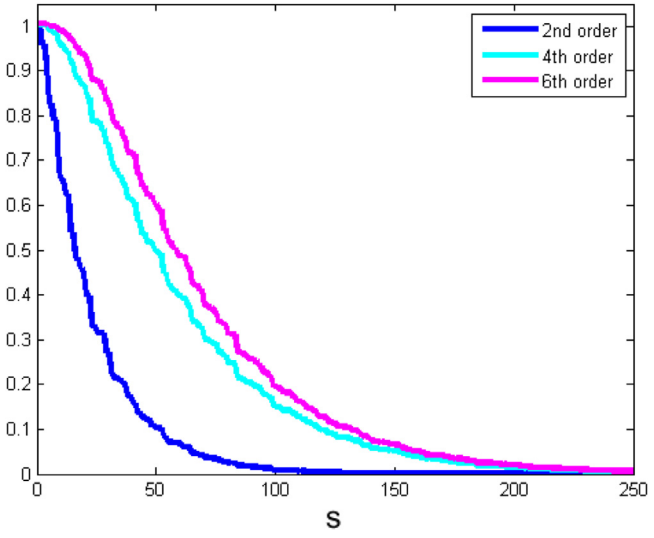

Fig. A.5. Spectral decay for second-order, fourth-order and sixth-order kernels using the synthetic data set. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6.2. Example 2

The second examples models high-dimensional locational based marginal prices (LBMP) of the day-ahead electricity market in the New York area. The time series is governed by a set of intrinsic parameters. These parameters capture physical components such as the season and economical factors, which may depend on the fuel prices and the stability of the market. Electricity prices are typically treated as a one-dimensional time series and modeled by parametric models such as ARIMA processes [13,20]. In [10], the authors explored manifold learning with local linear embedding for modeling electricity prices. In their model, the daily profile, which consists of 24 hourly price values is considered as a high dimensional data point. LLE then organizes the high-dimensional observed data into an embedded space by using the first top embedding coordinates. A low-dimensional organization can also be obtained by using diffusion maps [24].

The data for this example was downloaded from the NYISO website (www.nyiso.com), which holds data of 15 different zones in New York. This example uses 6 years of daily price curves from *Capital* zone between the dates 1.4.09–18.4.15. The observed data is treated as a high dimensional time series, where each point resides in $\mathbb{R}^{24}$ and holds a daily price profile. Fig. A.6 displays examples of three daily profiles. In this work, we first form small time-sequences by concatenating the data from three sequential days. This is done in an overlapping manner and results in a dataset of dimension 72. Our task is to predict the average value of day-ahead LBMP electricity profile. This is a one-dimensional function.

The training set that is denoted by $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_t\}$. The points $\mathbf{x}_t$ consist of 3-day profiles. For each training point, the average value of the day-ahead profile is calculated. This defines a function $f : \mathbb{R}^{72} \to \mathbb{R}$ on the train-data. The function f is known for all points in X. Given a new data point $\mathbf{y}$, which is a 3-day concatenation vector that ends with today's profile, the task is to predict the value of $f(\mathbf{y})$.

Please cite this article as: N. Rabin, D. Fishelov, Multi-scale kernels for Nyström based extension schemes, Applied Mathematics and Computation (2017), http://dx.doi.org/10.1016/j.amc.2017.02.025

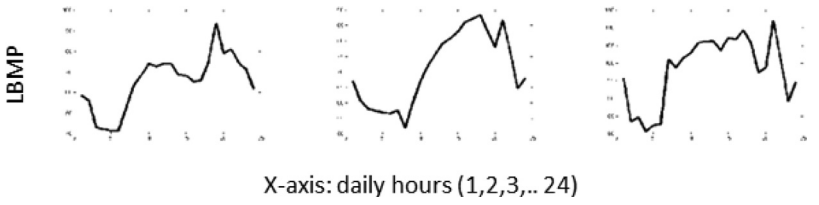

Fig. A.6. Three daily electricity profiles.

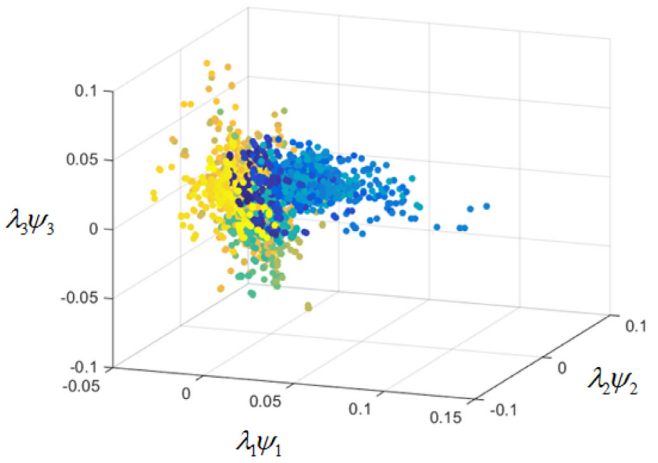

Fig. A.7. Diffusion maps embedding of 3-day electricity profiles.

The dataset X is embedded into a three-dimensional space by the application of diffusion maps (Algorithm 1). This results in a set of embedding coordinates $\Psi(X) = \{\lambda_1\psi_1(\mathbf{x}_j), \lambda_2\psi_2(\mathbf{x}_j), \lambda_3\psi_3(\mathbf{x}_j)\}_{j=1}^{t}$. The embedding is presented in Fig. A.7, where the points are colored by the months of the year. It can be seen that there is a typical behavior that corresponds to the season.

In order to predict the value of $f(\mathbf{y})$ for a new point $\mathbf{y}$, the constructed low-dimensional diffusion embedding $\Psi(X)$ is extended to include the new point by application of geometric harmonics (see Algorithm 2). The geometric harmonics are evoked twice, once with the standard Gaussian kernel and once with a fourth order kernel. The same condition number η is used. Once the value $\Psi(\mathbf{y})$ is calculated, we search for the K-nearest neighbors of $\Psi(\mathbf{y})$. The value of the function f is known on the training points. We evaluate the average day ahead price $f(\mathbf{y})$ by averaging the value of the function on its K-nearest neighbors.

One-hundred day-ahead average values were predicted (one at a time). The results were also compared with the application of linear PCA projection (instead of the diffusion maps embedding). In PCA, a straight forward linear extension of the principal component coordinates is used instead of the geometric harmonics.

Fig. A.8 presents the true average day-ahead curve (in black), the predicted result by applying geometric harmonics with a Gaussian kernel (in blue) and with a fourth-order kernel (in light-blue) and by using PCA (in red). Extending the embedding coordinates with the fourth-order kernel resulted in a more precise approximation, thus leading to better prediction of the day-ahead function. The root mean square error when using the Gaussian kernel in geometric harmonics was 8.586, the maximum error was 19.340. With the fourth-order kernel the root mean square error was 6.857 and the maximum error was 11.468. The root mean square error when using PCA was 12.217, the maximum error was 33.189.

Fig. A.9 shows a zoom-in of the first 50 predicted points.

It can be seen that the high-order kernel preserved more details and were able to more accurately reconstruct the predicted function.

7. Conclusion

In this paper we presented a family of high-order kernels for data embedding and extension. These kernels are easily constructed as linear combinations of scaled Gaussians. In addition, their spectral properties assure that the leading modes of the spectrum capture more information compared to the Gaussian kernel. Experimental results demonstrate the kernels' ability to capture finer features of the data. This results in a more accurate learning process.

Please cite this article as: N. Rabin, D. Fishelov, Multi-scale kernels for Nyström based extension schemes, Applied Mathematics and Computation (2017), http://dx.doi.org/10.1016/j.amc.2017.02.025

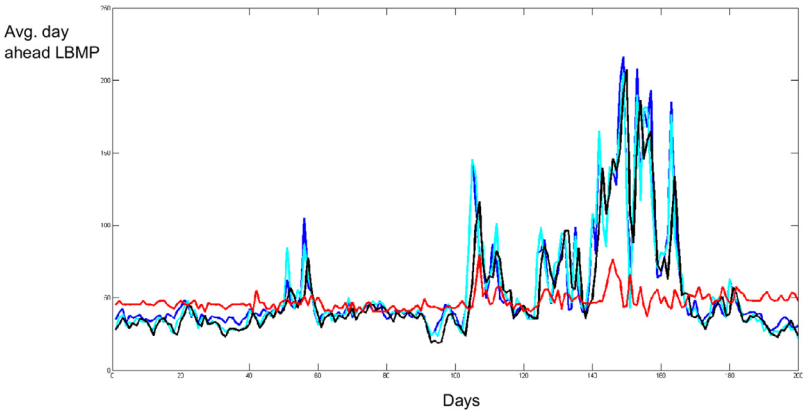

Fig. A.8. Day ahead average value prediction by using diffusion maps followed by geometric harmonics with a Gaussian kernel (in dark blue) and a fourth order kernel (in light blue). In red: prediction using PCA. In black: true day ahead value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

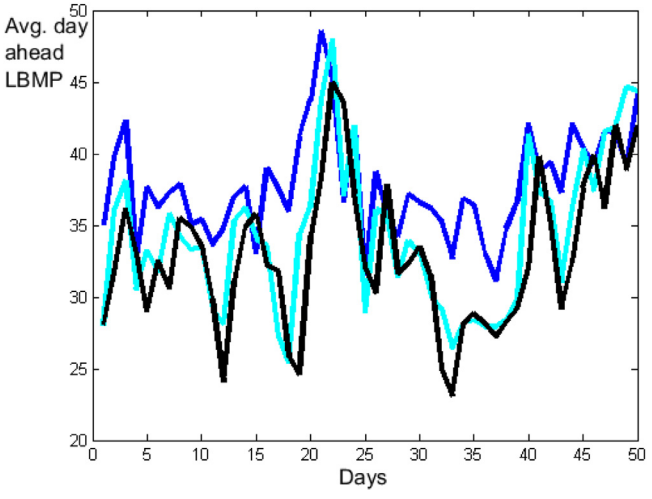

Fig. A.9. Prediction of 50 average day-ahead values by using diffusion maps followed by geometric harmonics with a Gaussian kernel (in dark blue) and a fourth order kernel (in light blue). In black: true day ahead value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Acknowledgment

This research was supported by a research grant from Afeka - Tel Aviv College of Engineering. We would like to thank the reviewers for their insightful comments on the paper.

Appendix A. High-order approximations

Lemma A.1. *Define* $k_{2,\delta}(r) = \frac{1}{\delta^2} k_2(r/\delta)$. *Let* $f : X \to \mathbb{R}$, *where* $X \subset \mathbb{R}^2$. *Assume that* $f^{(2)} \in L_2$, *where* $f^{(2)}$ *denotes all second-order derivatives of f. Then,*

$$\|f * k_{2,\delta} - f\|_{L^2} \leq C \|f^{(2)}\|_{L^2} \delta^2. \tag{A.1}$$

Proof. Using the and Plancherel equality, we have

$$\|f * k_{2,\delta} - f\|_{L^2} = \|\hat{f} \cdot (\hat{k}_{2,\delta} - 1)\|_{L^2}. \tag{A.2}$$

Let $\hat{k}_{2,\delta}$ be the Fourier transform of $k_{2,\delta}$, i.e.,

$$\hat{k}_{2,\delta}(\omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k_{2,\delta}(r) e^{i(\omega_1 x_1 + \omega_2 x_2)} dx_1 dx_2,$$

where $\omega = (\omega_1, \omega_2)$. Using Eq. (5), it follows that

$$\begin{aligned}
&\hat{k}_{2,\delta}(\omega = 0) = 1,\\
&\frac{\partial}{\partial \omega_i}\hat{k}_{2,\delta}(\omega = 0) = 0, \quad i = 1, 2,\\
&\left|\frac{\partial^2}{\partial \omega_i \partial \omega_j}\hat{k}_{2,\delta}(\omega)\right| \le C\delta^2 \quad i + j = 2.
\end{aligned} \tag{A.3}$$

Taylor expansion of $\hat{k}_{2,\delta}(\omega)$ around $\omega = 0$ yields

$$|\hat{k}_{2,\delta}(\omega) - 1| \le C\delta^2|\omega|^2. \tag{A.4}$$

Combining (A.2) with (A.4) results in

$$\|f * k_{2,\delta} - f\|_{L^2} \le C\|f^{(2)}\|_{L^2}\delta^2,$$

where $f^{(\alpha)} = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2}}$ and $|\alpha| = \alpha_1 + \alpha_2$. Eq. (A.1) shows that $k_{2,\delta}$ is a second-order kernel. □

References

[1] C. Baker, The Numerical Treatment of Integral Equations, Clarendon Press, Oxford, 1977.
[2] J.T. Beale, A. Majda, High order accurate vortex methods with explicit velocity kernels, J. Comput. Phys. 58 (1985) 188–208.
[3] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (2003) 1373–1396.
[4] M. Belkin, P. Niyogi, Semi-supervised learning on Riemannian manifolds, Mach. Learn. 56 (2004) 209–239.
[5] Y. Bengio, J.F. Paiement, P. Vincent, Out-of-sample extensions for LLE, isomap, MDS, eigenmaps, and spectral clustering, Université de Montréal, 2003 Technical report 1238.
[6] D. Comeau, Z. Zhao, D. Giannakis, A.J. Majda, Data-driven prediction strategies for low-frequency patterns of north pacific climate variability, Climate Dyn. 48 (5) (2017) 1855–1872.
[7] R.R. Coifman, S. Lafon, Diffusion maps, Appl. Comput. Harmonic Anal. 21 (2006a) 5–30.
[8] R.R. Coifman, S. Lafon, Geometric harmonics: a novel tool for multiscale out-of-sample extension of empirical functions, Appl. Comput. Harmonic Anal. 21 (2006b) 31–52.
[9] T. Cox, M. Cox, Multidimensional Scaling, Chapman and Hall, London, UK, 1994.
[10] S. Deng, J. Chen, X. Huo, Electricity price curve modeling and forecasting by manifold learning, IEEE Trans. Power Syst. 23 (2008) 877–888.
[11] D.L. Donoho, C. Grimes, Hessian eigenmaps: new locally linear embedding techniques for high dimensional data, Proc. Natl. Acad. Sci. 100 (2003) 5591–5596.
[12] D. Duncan, R. Talmon, H.P. Zaveri, R.R. Coifman, Identifying preseizure state in intracranial EEG data using diffusion kernels, Spec. Issue Math. Biosci. Eng. (MBE) 10 (2013) 579–590.
[13] R. Espnola, A.J. Conejo, M.A. Plazas, A.B. Molina, Day-ahead electricity price forecasting using the wavelet transform and ARIMA models, IEEE Trans. Power Syst. 20 (2015) 1035–1042.
[14] A. Fernndez, N. Rabin, R.R. Coifman, J. Eckstein, Diffusion methods for aligning medical datasets: location prediction in CT scan images, Med. Image Anal. 18 (2013) 425–432.
[15] D. Fishelov, A new vortex scheme for viscous flows, J. Comput. Phys. 86 (1990) 211–244.
[16] G. Gilboa, N. Sochen, Y. Zeevi, Forward-and-backward diffusion processes for adaptive image enhancement and denoising, IEEE Trans. Image Process. 11 (2002) 689–703.
[17] O.H. Hald, Convergence of vortex methods for Eulers equations III, SIAM J. Numer. Anal. 24 (1987) 538–582.
[18] Y. Keller, R.R. Coifman, S. Lafon, S.W. Zucker, Audio-visual group recognition using diffusion maps, IEEE Trans. Signal Process. 58 (2010) 403–413.
[19] S. Lafon, Y. Keller, R.R. Coifman, Data fusion and multicue data matching by diffusion maps, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 1784–1797.
[20] F.J. Nogales, J. Contreras, R. Espnola, A.J. Conejo, Arima models to predict next-day electricity prices, IEEE Trans. Power Syst. 18 (2003) 1014–1020.
[21] H.O. Nordmark, Rezoning for high order vortex methods, J. Comput. Phys. 97 (1991) 366–397.
[22] D. Pleptin, H.O. Pollak, Prolate spheroidal wave functions, Fourier analysis and uncertainty I, Bell Syst. Tech. J. 40 (1961) 43–64.
[23] N. Rabin, A. Averbuch, Detection of anomaly trends in dynamically evolving systems, AAAI Fall Symp. Ser. (2010) 44–49.
[24] N. Rabin, R.R. Coifman, Modeling zonal electricity prices by anisotropic diffusion embeddings, Yale University Research Report #1457, 2012.
[25] N. Rabin, R.R. Coifman, Heterogeneous datasets representation and learning using diffusion maps and Laplacian pyramids, Proceedings of the 12th SIAM International Conference on Data Mining (SDM-2012), Anaheim, CA, USA, 2012, pp. 189–199.
[26] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000) 2323–2326.
[27] A. Schclar, A diffusion framework for dimensionality reduction, in: Soft Computing for Knowledge Discovery and Data Mining, Springer, US, 2008, pp. 315–325. ISBN: 978-0-387-69934-9.
[28] A. Schclar, A. Averbuch, N. Rabin, V. Zheludev, K. Hochman, A diffusion framework for detection of moving vehicles, Digital Signal Process. 20 (2010) 111–122.
[29] A. Singer, Y. Shkolnisky, B. Nadler, Diffusion interpretation of nonlocal neighborhood filters for signal denoising, SIAM J. Imag. Sci. 2 (2009) 118–139.
[30] K.P. Stanton, F. Parisi, F. Strino, N. Rabin, P. Asp, Y. Kluger, Arpeggio: harmonic compression of chIP-seq data reveals protein-chromatin interaction signatures, Nucleic Acids Res. 41 (2013) e161.
[31] R. Talmon, I. Cohen, S. Gannot, Single-channel transient interference suppression with diffusion maps, IEEE Transa. Audio, Speech, Lang. Process. 21 (2013) 132–144.
[32] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemom. Intell. Lab. Syst. 2 (1987) 37–52.
[33] G. Wolf, A. Averbuch, P. Neittaanmäki, Parameter rating by diffusion gradient, in: Modeling, Simulation and Optimization for Science and Technology, in: The Series Computational Methods in Applied Sciences, vol. 34, 2014, pp. 225–248.
[34] Z. Zhang, H. Zha, Principal manifolds and nonlinear dimension reduction via local tangent space alignment, Department of Computer Science and Engineering, Pennsylvania State University, 2002 Technical report cse-02-019.

Please cite this article as: N. Rabin, D. Fishelov, Multi-scale kernels for Nyström based extension schemes, Applied Mathematics and Computation (2017), http://dx.doi.org/10.1016/j.amc.2017.02.025