MULTIPLE KERNEL INTERPOLATION FOR INVERTING NON-LINEAR DIMENSIONALITY REDUCTION AND DIMENSION ESTIMATION

Jayaraman J. Thiagarajan and Peer-Timo Bremer

Lawrence Livermore National Laboratory
{jayaramanthi1, bremer5}@llnl.gov

Karthikeyan Natesan Ramamurthy

IBM Thomas J. Watson Research Center
knatesa@us.ibm.com

ABSTRACT

The problem of stably inverting a non-linear dimensionality reduction map has applications in data visualization and machine learning, besides being of theoretical interest. In this paper, we propose a meshfree interpolation method for obtaining such inverse maps using a non-negative linear combination of multiple interpolants. We show that the proposed scheme can improve upon the approximation power of its individual constituent kernels, and discuss the conditions under which its parameters can be uniquely estimated. We also provide an approach for estimating the intrinsic dimensionality (ID) of manifolds using the proposed inverse map. Experiments using multiple kernel interpolation for reconstruction of novel test data and ID estimation show an improved or similar performance compared to existing techniques.

Index Terms— manifold learning, inverse map, kernel interpolation, intrinsic dimension estimation

1. INTRODUCTION

The high-dimensional data generated by many natural and artificial processes usually reside in a manifold of low intrinsic dimensionality. A variety of non-linear dimensionality reduction (NLDR) methods have been proposed to infer the intrinsic structure [1, 2, 3]. Let us consider a bounded smooth manifold $\mathcal{M} \subset \mathbb{R}^D$ of low intrinsic dimensionality, and assume that we have a set of P samples $\{\mathbf{x}_j\}_{j=1}^P$ from $\mathcal{M}$. Using an appropriate non-linear mapping, we embed the data samples in $\mathbb{R}^d$ using the map $\Phi : \mathbb{R}^D \mapsto \mathbb{R}^d$ to obtain $\{\mathbf{y}_j\}_{j=1}^P$, where $d < D$. We also define $\mathbf{X} \in \mathbb{R}^{D\times P}$ and $\mathbf{Y} \in \mathbb{R}^{d\times P}$ as the matrices containing the respective samples. Obtaining a stable estimate of the inverse map, $\Phi^\dagger : \mathbb{R}^d \mapsto \mathbb{R}^D$, will allow us to reconstruct the high-dimensional data samples.

A good inverse map should preserve the local neighborhood information, reproduce the curvature of the manifold, and provide a reasonable recovery for samples near the boundary. A linear inverse map of the form $\hat{\Phi}^\dagger(\mathbf{y}) = \sum_{j:\mathbf{y}_j\in N_\mathbf{y}} c_j\mathbf{x}_j$ considered in [4], preserves the neighborhood information only for points away from the boundary. Inverse maps obtained using radial basis function (RBF) kernels are more effective in reconstructing the curvature and providing a good approximation near the boundary [5].

In this paper, we propose a method for reconstructing high-dimensional samples in $\mathcal{M}$ using a non-negative linear combination of multiple interpolation kernels. When compared to the RBF-based approaches in [5, 6], we show that the proposed multiple kernel approach can provide a good approximation for a larger set of functions. We also experimentally demonstrate the superior recovery performance of the proposed scheme in comparison to the RBF-based schemes. Furthermore, we develop an approach to estimate the intrinsic dimensionality (ID) of a manifold, by analyzing the quality of interpolation for various embedding dimensions. ID estimation is affected by the sampling of the manifold as well as the presence of noise. Using synthetic and real datasets, we show that our proposed scheme is robust to these issues, and compares well with state-of-the-art approaches for intrinsic dimension estimation [7, 8, 9].

2. INVERSE MAPPING USING INTERPOLATION

We provide an overview of RBF interpolation and its use in computing the inverse map. An RBF interpolant of a sample $\mathbf{y}$ is given by a linear combination of radially symmetric kernel functions, $f(\mathbf{y}) = \sum_{j=1}^P c_j(\mathbf{y})k(\mathbf{y}, \mathbf{y}_j)$ where $k(\mathbf{y}, \mathbf{y}_j) = g(\|\mathbf{y} - \mathbf{y}_j\|_2)$ for some appropriate function $g(.)$, and $c_j(\mathbf{y})$ are the interpolation coefficients. Given $\mathbf{X} \in \mathbb{R}^{D\times P}$ and the corresponding embedding $\mathbf{Y} \in \mathbb{R}^{d\times P}$ the strategy is to find a unique interpolation function for each dimension in $\mathbb{R}^D$. Therefore, the interpolation coefficient matrix $\mathbf{C} \in \mathbb{R}^{P\times D}$ can be determined by solving $\mathbf{KC} = \mathbf{X}^T$ where $\mathbf{K} \in \mathbb{R}^{P\times P}$ is the kernel matrix whose (i, j) element is $k(\mathbf{y}_i, \mathbf{y}_j)$, the columns of $\mathbf{X} \in \mathbb{R}^{D\times P}$ denote the samples from $\mathcal{M}$, and each column of $\mathbf{C} \in \mathbb{R}^{P\times D}$ represents the unique interpolation coefficients for that dimension. For a novel sample $\mathbf{y}$, the $D-$dimensional sample $\hat{\mathbf{x}}$ is estimated as $\Phi^\dagger(\mathbf{y}) = \mathbf{C}^T[k(\mathbf{y}, \mathbf{y}_1) \cdots k(\mathbf{y}, \mathbf{y}_P)]^T$.

In order for the system $\mathbf{KC} = \mathbf{X}^T$ to be uniquely solv-

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-CONF-645866

able, the kernel matrix $\mathbf{K}$ needs to be non-singular and this is true for strictly positive definite (SPD) kernels. The Gaussian RBF kernel $k(\mathbf{y}_i, \mathbf{y}_j) = \exp(-\epsilon\|\mathbf{y}_i - \mathbf{y}_j\|_2)$, where ϵ is the scale parameter, is an example of an SPD kernel. Since it is prone to numerical ill-conditioning at smaller scales when the density of the data samples is low, the authors in [5] propose to use scale-free RBF kernels for computing the inverse mapping. Examples of these are the cubic RBF defined as $k(\mathbf{y}_i, \mathbf{y}_j) = \|\mathbf{y}_i - \mathbf{y}_j\|_2^3$, and the 2nd order thin plate splines defined as $k(\mathbf{y}_i, \mathbf{y}_j) = \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 \log \|\mathbf{y}_i - \mathbf{y}_j\|_2$. The performance of the cubic RBF near the boundary can be improved by modifying the interpolant as

$$f(\mathbf{y}) = \sum_{j=1}^{P} c_j(\mathbf{y}) k(\mathbf{y}, \mathbf{y}_j) + \mathbf{y}^T \boldsymbol{\alpha}_1 + \alpha_0$$
$$\text{subj. to } \sum_{j=1}^{P} c_j = 0, \sum_{j=1}^{P} c_j \mathbf{y}_j = 0, \qquad (1)$$

where $\boldsymbol{\alpha}_0$ is a constant term and $\boldsymbol{\alpha}_1 \in \mathbb{R}^d$, and this new form is known as the natural cubic interpolant. We will now state the conditions under which the parameters can be uniquely inferred for an important class of kernels [5, 6, 10].

Definition A real valued continuous even function $k : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$ is called conditionally positive definite (CPD) of order m on $\mathbb{R}^d$ if $\sum_{i=1}^{P}\sum_{j=1}^{P} a_i a_j k(\mathbf{y}_i, \mathbf{y}_j) \geq 0$, for the set of P points in $\mathbb{R}^d$ subject to $\sum_{j=1}^{P} a_j \rho(\mathbf{y}_i) = 0$, where $\rho(\mathbf{y})$ is any real-valued polynomial of degree at most $m-1$. If in addition, the quadratic form is zero if and only if $a_i = 0, \forall i$, then k is strictly conditionally positive definite (SCPD).

Theorem 2.1. *[6] If the real valued continuous function $k : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$ is SCPD of order m and the points $\{\mathbf{y}_1, \cdots, \mathbf{y}_P\}$ form an $(m-1)-$ unisolvent set, then the following system is uniquely solvable:*

$$\begin{bmatrix} \mathbf{K} & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}. \qquad (2)$$

Here, $\mathbf{P}$ is the polynomial basis matrix with (j,n) element given by $\rho_n(\mathbf{y}_j)$ for $j = \{1, \ldots, P\}$ and $n = \{1, \ldots, M\}$. $\{\rho_n(\mathbf{y})\}_{n=1}^{M}$ forms the basis for the linear space of all polynomials up to degree $(m-1)$, and $\mathbf{f}$ is a vector in $\mathbb{R}^P$.

Note that a set of points $\{\mathbf{y}_j\}_{j=1}^{P}$ is $m-$unisolvent if the only polynomial of total degree at most m, interpolating zero data on the set is the zero polynomial. Note that both the cubic RBF and the 2nd order thin plate spline are SCPD of order 2 with $\mathbf{P} = [\mathbf{1}\ \mathbf{Y}]^T$ and $\boldsymbol{\alpha} = [\alpha_0\ \boldsymbol{\alpha}_1^T]^T$.

3. PROPOSED MULTIPLE KERNEL INTERPOLATION

Although the idea of using a single interpolation kernel is mathematically sound, the class of functions that can be approximated using each kernel is limited. Therefore, we propose to use a non-negative linear combination of L different kernels. The l^{th} interpolant f_l can be written using its respective kernel k_l as $f_l(\mathbf{y}) = \sum_{j=1}^{P} c_j(\mathbf{y}) k_l(\mathbf{y}, \mathbf{y}_j)$. It is clear that f_l lies in the linear span of $\{k_l(., \mathbf{y}_j)\}_{j=1}^{P}$, and the coefficients of the linear combination are $\{c_j\}_{j=1}^{P}$. The native space $\mathcal{N}_{k_l}$ of the kernel k_l is defined as the completion of its span, and clearly any kernel can only approximate functions in its native space [10]. For a Gaussian RBF, the native space is not very large, since it is restricted to functions whose Fourier transform decays at least as fast as the Gaussian [6].

The proposed multiple kernel interpolant (MKI) is $f(\mathbf{y}) = \sum_{l=1}^{K} \beta_l f_l(\mathbf{y})$ where $\beta_l \geq 0, l = \{1, \ldots, L\}$ are the kernel weights. We will investigate the proposed MKI by addressing these two questions: (a) can the multiple kernel interpolant provide a better approximation of functions compared to its constituents? and, (b) is it possible to learn the interpolant coefficients uniquely from the data?

Lemma 3.1. *Given L kernels $\{k_l\}_{l=1}^{L}$, the native space of the ensemble kernel $k = \sum_l \beta_l k_l$ is at least as big as its constituent kernels, where the kernel weights $\beta_l \in \mathbb{R}^+, \forall l$.*

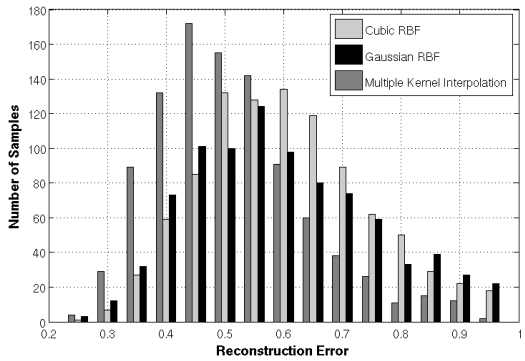
Proof. Since any function that can be approximated by the ensemble kernel at a point $\mathbf{y}$ can be written as

$$f(\mathbf{y}) = \sum_{l=1}^{L} \beta_l \sum_{j=1}^{P} c_j(\mathbf{y}) k_l(\mathbf{y}, \mathbf{y}_j), \qquad (3)$$

where we assume the kernels to be defined over the discrete domain $\{\mathbf{y}_j\}_{j=1}^{P}$. Clearly for each set of coefficients $\{c_j(\mathbf{y})\}_{j=1}^{P}$, the final approximation lies in the convex cone of the individual approximations. Hence the native space of the ensemble kernel is the union of such cones, which is at least as big as the individual native spaces of kernels. If the native space of one kernel is not contained within that of the other, the native space of the ensemble is bigger. □

As an example, let us consider an ensemble of Gaussian RBF and cubic RBF kernels. Since the native spaces of these two kernels are different [10], using the ensemble kernel can provide us with a better approximation of a function. However, we need to ensure that the coefficients for the ensemble kernel can be uniquely learned using training data. We define the ensemble kernel matrix as $\mathbf{K} = \sum_{l=1}^{L} \beta_l \mathbf{K}_l$ where the $(i,j)^{\text{th}}$ entry of $\mathbf{K}_l$ is $k_l(\mathbf{y}_i, \mathbf{y}_j)$. We assume that each kernel in the ensemble is strictly conditionally positive definite (SCPD) with a maximum order of m, or strictly positive definite (SPD). When the function to be approximated lies in the native space of the ensemble kernel, the following lemma provides conditions under which the ensemble kernel coefficients can be uniquely obtained.

Lemma 3.2. *If the kernels $\{\mathbf{K}_l\}_{l=1}^{L}$ are SCPD of order at most m or SPD, and $\{\mathbf{y}_j\}_{j=1}^{P}$ form aa $(m-1)$-unisolvent*



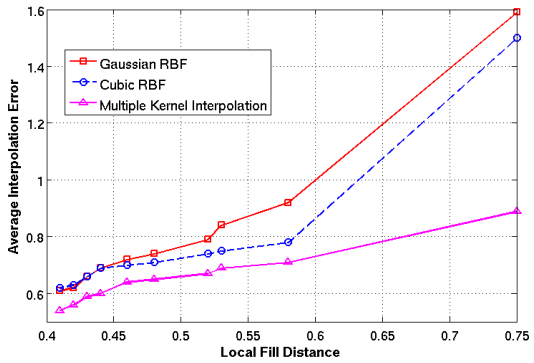


Fig. 1. The histogram of reconstruction errors (left), and the performance of the algorithms with increasing fill distance (right).

Table 1. Comparison of average interpolation errors obtained for the MNIST digit dataset using different kernels.

Digit	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Gaussian RBF	0.61	0.45	0.73	0.78	0.73	0.75	0.69	0.70	0.75	0.71
Cubic RBF	0.62	0.41	0.70	0.71	0.67	0.69	0.63	0.63	0.71	0.64
Multiple Kernel	**0.54**	**0.35**	**0.64**	**0.64**	**0.60**	**0.61**	**0.56**	**0.54**	**0.64**	**0.55**

set,

$$\begin{pmatrix} \sum_{l=1}^{L} \beta_l \mathbf{K}_l & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{c} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{0} \end{pmatrix} \tag{4}$$

has a unique solution if $\beta_l \geq 0, l = \{1, \ldots, L\}$. *Here,* $\mathbf{P}$ *is the polynomial basis matrix with* (j, n) *element given by* $\rho_n(\mathbf{y}_j)$ *for* $j = \{1, \ldots, P\}$ *and* $n = \{1, \ldots, M\}$. $\{\rho_n(\mathbf{y})\}_{n=1}^{M}$ *forms the basis for the linear space of all polynomials up to degree* $(m-1)$, *and* $\mathbf{f}$ *is a vector in* $\mathbb{R}^P$.

Proof. In order to show that (4) is uniquely solvable, we must show that $\mathbf{0}$ is the only element in the null space of the LHS matrix in (4). Considering the first row block of the matrix, we have $\sum_{l=1}^{L} \beta_l \mathbf{K}_l \mathbf{c} + \mathbf{P}\boldsymbol{\alpha} = \mathbf{0}$. Pre-multiplying by $\mathbf{c}^T$, we have $\sum_{l=1}^{L} \beta_l \mathbf{c}^T \mathbf{K}_l \mathbf{c} + \mathbf{c}^T \mathbf{P}\boldsymbol{\alpha} = \mathbf{0}$. Since $\mathbf{c}^T \mathbf{P} = \mathbf{0}^T$ from the bottom block of (4), we have $\sum_{l=1}^{L} \beta_l \mathbf{c}^T \mathbf{K}_l \mathbf{c} = 0$. This is true if and only if $\mathbf{c} = \mathbf{0}$, since each kernel is either SCPD of order at most m, or SPD and a positive linear combination is zero only if each of the elements are zero. If a particular β_l is zero, that kernel is not used in the approximation. Now, since $\mathbf{P}\boldsymbol{\alpha} = \mathbf{0}$, the $(m-1)$ unisolvency condition enforces that $\boldsymbol{\alpha} = \mathbf{0}$ and hence the solution to (4) is unique. □

Algorithm: The ensemble kernel coefficients and the kernel weights can be inferred from data by solving

$$\{\hat{\mathbf{C}}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{A}}\} = \underset{\{\mathbf{C}, \boldsymbol{\beta}, \mathbf{A}\}}{\operatorname{argmin}} \|\mathbf{X}^T - \sum_{l=1}^{L} \beta_l \mathbf{K}_l \mathbf{C} - \mathbf{P}\mathbf{A}\|_F^2 + \lambda_1 \|\boldsymbol{\beta}\|_1 + \lambda_2 \|\boldsymbol{\beta}\|_2^2 \quad \text{subj. to } \mathbf{P}^T \mathbf{C} = \mathbf{0}, \tag{5}$$

where $\boldsymbol{\beta} = [\beta_l]_{l=1}^{L}$ and the coefficient matrices are $\mathbf{C} = [\mathbf{c}_i]_{i=1}^{D}$ and $\mathbf{A} = [\boldsymbol{\alpha}_i]_{i=1}^{D}$. The squared error loss function ensures a high fidelity interpolation. The elastic net penalty, which is a combination of the ℓ_1 and ℓ_2 penalties regularizes the problem by ensuring that important kernels will be chosen for interpolation and similar kernels will be chosen together as a group. Eqn. (5) is non-convex jointly and hence can be solved as an alternating minimization. We fix $\boldsymbol{\beta}$ to compute $\{\mathbf{C}, \mathbf{A}\}$ by solving the linear system in (4), which has a unique solution, and then estimate $\boldsymbol{\beta}$ using penalized least squares, while fixing the other two parameters. The penalties λ_1 and λ_2 are tuned using a cross validation procedure.

Performance Evaluation: To evaluate the proposed multiple kernel interpolation approach, we consider a set of handwritten digits from the MNIST database [11]. The dataset consisted of 1000 randomly chosen samples from each digit $(0-9)$. Similar to the procedure in [5], the images were resized to 14×14, reshaped into vectors, centered and normalized to unit Euclidean norm. They were then projected to a $10-$dimensional space using Laplacian Eigenmaps, with the number of neighbors fixed at 15. The inverse mapping was evaluated for all samples in the set and compared to the original to measure the interpolation error (ℓ_2 distance). We used $L = 10$ kernels, (a) cubic, (b) thin plate spline, and (c) 8 Gaussians with $\epsilon \in \{0.01, 0.05, 0.1, 0.5, 1, 10, 100, 1000\}$. Figure 1(left) shows the histogram of the reconstruction errors for Digit0 obtained using the cubic, best performing Gaussian and the proposed MKI. The average errors for the different digits are reported in Table 1. By combining the multiple kernels, our method provides improved recovery when compared to the individual kernels.

Convergence of RBF interpolants can be studied with respect to the local fill distance, which provides an ap-

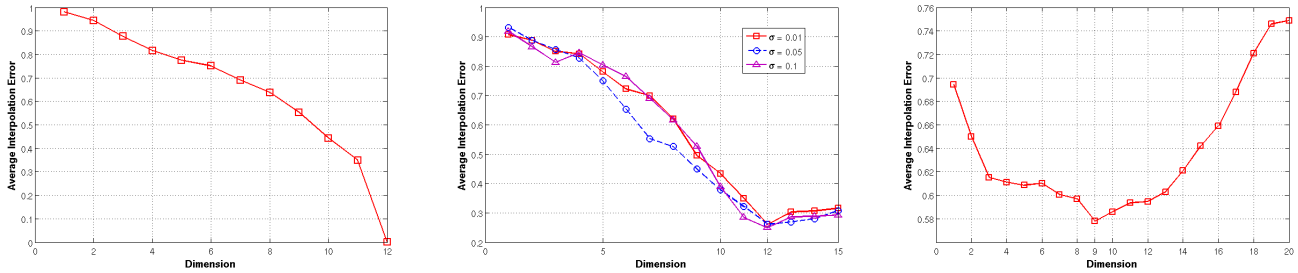


Fig. 2. Intrinsic dimension estimation - $P(d, r)$ vs d for different datasets: $\mathbb{Q}^{12}$ embedded in 100D (left), $\mathbb{Q}^{12}$ corrupted by additive Gaussian noise (middle), sciencenews data embedded in 200D (right).

Table 2. Comparison of the dimension estimation performance on different synthetic and real datasets.

Dataset	MLE	kNN	MSVD	Proposed
$\mathbb{Q}^6$	5	6	6	6
$\mathbb{Q}^{12}$	9	12	12	12
$\mathbb{Q}^{48}$	25	32	48	48
$\mathbb{S}^{11}$	9	11	11	11
$\mathbb{S}^{23}$	16	18	23	24
$\mathcal{S}$	2	2	2	2
Isomap faces	4.3	-	2	2
Science news	11	-	9	9
Face videos	6.6	-	2	3

proximate measurement of node spacing under randomized sampling schemes. The local fill distance is measured as the average distance to the nearest neighbor, $h_{loc} = \frac{1}{P}\sum_{i=1}^{P} \min_{j \neq i} \|\mathbf{x}_i - \mathbf{x}_j\|_2$. We used multiple random subsets of Digit0, with number of samples varied between 100 and 1000. In each case we computed the local fill distance and the average interpolation errors. Figure 1(right) illustrates the convergence of the different interpolation schemes as a function of h_{loc}, where improvements obtained with the MKI scheme is evident.

4. INTRINSIC DIMENSION ESTIMATION

In the proposed MKI approach, the quality of inversion for novel samples from $\mathbb{R}^d$ depends on the embedding dimension (d). If d matches the intrinsic dimension of the manifold, we can expect that the reconstruction to be accurate. We propose to use the fidelity of the inverse mapping as an indicator for estimating the intrinsic dimension.

The quality of the inverse map at a neighborhood can be measured using the weighted mean of the reconstruction errors for a sample and its neighbors. For any sample $\mathbf{x}_i$, we consider the corresponding $\mathbf{y}_i$ in the embedding as a novel sample and obtain $\hat{\mathbf{x}}_i = \Phi^{\dagger}(\mathbf{y}_i)$. The interpolation error for $\mathbf{x}_i$ is measured as $E_i = \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2$. For a neighborhood of radius r centered at $\mathbf{x}_i$, the average interpolation error is $\overline{E}_i = \sum_{j \in \{i, N_i\}} w_j E_j$, where $w_j = \exp(-\epsilon \|\mathbf{x}_i - \mathbf{x}_j\|_2)$. The parameter ϵ is fixed as the inverse of the mean distance between $\mathbf{x}_k$ and its neighbors. We compute the interpolation error for a random subset of samples (P_s) in the data, and compute $P(d, r) = \frac{1}{P_s}\sum_i \overline{E}_i$. We repeat this procedure for increasing values of d and determine the intrinsic dimension that provides the least interpolation error.

To evaluate the performance of the proposed approach in intrinsic dimension estimation, we considered a set of synthetic and real datasets: (a) k-dimensional cube, $\mathbb{Q}^k$, (b) k-dimensional sphere, $\mathbb{S}^k$, (c) manifold product of an S-shaped curve and a unit interval, (d) Isomap faces [12], (e) science news dataset [9], and (f) face videos [13]. The ambient dimensions and number of samples for (a), (b) and (c) were fixed at $D = 100$ and $P = 1000$ respectively. In each case, the radius r was chosen appropriately and the embeddings were computed using LLE for (a)-(c), ISOMAP [3] for (d), and Laplacian Eigenmaps for (e)-(f).

Table 2 shows the dimension estimates obtained for different datasets with our method in comparison to three existing techniques, maximum likelihood estimation (MLE) [7], kNN based estimation [8], and multiscale SVD (MSVD) [9]. Figure 2 shows the average error $P(d, r)$, at different values of d, for $\mathbb{Q}^{12}$ with no noise and with additive Gaussian noise. In cases where the MKI recovers the underlying map exactly, the interpolation error drops arbitrarily close to zero for an appropriate d (Figure 2). In the presence of noise, the samples lie outside the smooth manifold structure, and hence the reconstruction error will be non-zero even when d equals the ID. When the manifold is undersampled, the forward map to create an embedding is not accurate since it depends on the poorly sampled local neighborhood. Hence, the inverse map cannot produce an exact reconstruction for any d. Another interesting observation is that $P(d, r)$ increases as d is varied beyond the actual ID. A likely reason is that as d is increased, the neighbors chosen by the interpolation scheme in $\mathbb{R}^d$ might actually be far away (in geodesic distance) on $\mathcal{M}$ in the high dimensional space. As a result, $P(d, r)$ is distinctively low only for a very few choices of d, and there is a clear minimum unless there is a severe undersampling.

5. REFERENCES

[1] Sam T Roweis and Lawrence K Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[2] Mikhail Belkin and Partha Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[3] Joshua B Tenenbaum, Vin De Silva, and John C Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[4] Dan Kushnir, Ali Haddad, and Ronald R Coifman, "Anisotropic diffusion on sub-manifolds with application to earth structure classification," *Applied and Computational Harmonic Analysis*, vol. 32, no. 2, pp. 280–294, 2012.

[5] Nathan D Monnig, Bengt Fornberg, and Francois G Meyer, "Inverting non-linear dimensionality reduction with scale-free radial basis interpolation," *arXiv preprint arXiv:1305.0258*, 2013.

[6] Gregory E Fasshauer, *Meshfree approximation methods with MATLAB*, vol. 6, World Scientific, 2007.

[7] Elizaveta Levina and Peter J Bickel, "Maximum likelihood estimation of intrinsic dimension," in *Advances in neural information processing systems*, 2004, pp. 777–784.

[8] Kevin M Carter and Alfred O Hero, "Variance reduction with neighborhood smoothing for local intrinsic dimension estimation," in *IEEE ICASSP*, 2008, pp. 3917–3920.

[9] Anna V Little, Jason Lee, Yoon-Mo Jung, and Mauro Maggioni, "Estimation of intrinsic dimensionality of samples from noisy low-dimensional manifolds in high dimensions with multiscale SVD," in *Proc. IEEE SSP*, 2009, pp. 85–88.

[10] Holger Wendland, *Scattered data approximation*, vol. 17, Cambridge University Press, 2005.

[11] Yann Lecun and Corinna Cortes, "The MNIST database of handwritten digits," http://yann.lecun.com/exdb/mnist/.

[12] "Isomap Faces," http://isomap.stanford.edu/dataset.html.

[13] "Face Video," http://www.cs.toronto.edu/roweis/data.html.