Appl. Comput. Harmon. Anal. • • • (• • • •) • • •–• • •

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Iterated diffusion maps for feature identification

Tyrus Berry [a,*], John Harlim [b,c]

[a] *Department of Mathematical Sciences, George Mason University, 4400 Exploratory Hall, Fairfax, VA 22030, USA*
[b] *Department of Mathematics, the Pennsylvania State University, 214 McAllister Building, University Park, PA 16802, USA*
[c] *Department of Meteorology and Atmospheric Science, the Pennsylvania State University, 503 Walker Building, University Park, PA 16802, USA*

ARTICLE INFO

Article history:
Received 25 September 2015
Received in revised form 23 August 2016
Accepted 26 August 2016
Available online xxxx
Communicated by Charles K. Chui

Keywords:
Diffusion maps
Local kernel
Iterated diffusion map
Dimensionality reduction
Feature identification

ABSTRACT

Recently, the theory of diffusion maps was extended to a large class of *local kernels* with exponential decay which were shown to represent various Riemannian geometries on a data set sampled from a manifold embedded in Euclidean space. Moreover, local kernels were used to represent a diffeomorphism $\mathcal{H}$ between a data set and a feature of interest using an anisotropic kernel function, defined by a covariance matrix based on the local derivatives $D\mathcal{H}$. In this paper, we generalize the theory of local kernels to represent degenerate mappings where the intrinsic dimension of the data set is higher than the intrinsic dimension of the feature space. First, we present a rigorous method with asymptotic error bounds for estimating $D\mathcal{H}$ from the training data set and feature values. We then derive scaling laws for the singular values of the local linear structure of the data, which allows the identification the tangent space and improved estimation of the intrinsic dimension of the manifold and the bandwidth parameter of the diffusion maps algorithm. Using these numerical tools, our approach to feature identification is to iterate the diffusion map with appropriately chosen local kernels that emphasize the features of interest. We interpret the iterated diffusion map (IDM) as a discrete approximation to an intrinsic geometric flow which smoothly changes the geometry of the data space to emphasize the feature of interest. When the data lies on a manifold which is a product of the feature manifold with an irrelevant manifold, we show that the IDM converges to the quotient manifold which is isometric to the feature manifold, thereby eliminating the irrelevant dimensions. We will also demonstrate empirically that if we apply the IDM to features which are not a quotient of the data manifold, the algorithm identifies an intrinsically lower-dimensional set embedding of the data which better represents the features.

* Corresponding author.
E-mail addresses: tberry@gmu.edu (T. Berry), jharlim@psu.edu (J. Harlim).

http://dx.doi.org/10.1016/j.acha.2016.08.005
1063-5203/

1. Introduction

Often, for high-dimensional data and especially for data lying on a nonlinear subspace of Euclidean space, the variables of interest do not lie in the directions of largest variance and this makes them difficult to identify. The *features* (variables of interest) may be nonlinear functions of the ambient Euclidean coordinates. Moreover, other nonlinear combinations of the ambient coordinates may be independent of the variables of interest, and should be eliminated; we call these quantities the *irrelevant variables*. For example, consider the annulus shown in Fig. 1, where the feature of interest is the radius as indicated by the color. The feature of interest is a nonlinear function of the ambient coordinates, namely $r = \sqrt{x^2 + y^2}$, and is completely independent of the irrelevant variable $\theta = \tan^{-1}(y/x)$. We should mention that a related direction which is being explored in the current research attempts to discover features which are common in multiple 'views' [9,6,18] using cross-diffusion between views and nonlinear canonical correlation analysis [10]. In this paper, we will consider the case when the desired feature is known on a training data set and we wish to learn the feature map in order that it may be extended to new data points. In other words, we consider the supervised learning problem, that is, to learn the underlying map that takes the data space to the feature space using a training data set that includes the feature values. In particular, we are seeking a representation of the feature map which can be extended to new data points.

Throughout this manuscript we will assume that the training data set consists of data points which lie near a d-dimensional manifold $\mathcal{M} \subset \mathbb{R}^m$ embedded in an m-dimensional Euclidean space; we refer to $\mathcal{M}$ as the *data space* or *data manifold* and we refer to $\mathbb{R}^m$ as the *ambient data space*. We also assume that we have a set of feature values corresponding to each training data point, and these feature values are assumed to lie near a $d_{\mathcal{N}}$-dimensional manifold $\mathcal{N} \subset \mathbb{R}^n$ embedded in an n-dimensional Euclidean space; we refer to $\mathcal{N}$ as the *feature space* or *feature manifold* and we refer to $\mathbb{R}^n$ as the *ambient feature space*. We do not assume any knowledge of the structure of the manifolds $\mathcal{M}, \mathcal{N}$ or the feature map $\mathcal{H} : \mathcal{M} \to \mathcal{N}$, we only assume that the feature map is differentiable.

When the feature manifold is intrinsically lower-dimensional than the data manifold, the data manifold contains information which is irrelevant to the feature, and we refer to this information broadly as the 'irrelevant variables' or the 'irrelevant space'. In some contexts it is possible to identify the irrelevant space explicitly, for example the data manifold may simply be a product manifold of the feature manifold and an irrelevant manifold. This is exactly the case with the annulus, which is a product manifold of the feature space $[0, 1] \ni r$ with the irrelevant space $[0, 2\pi) \ni \theta$. However, more complex relationships between the data manifold, feature manifold, and irrelevant variables are possible.

In this paper, we generalize a method introduced in [4], which was developed for representing diffeomorphisms to more general maps which are differentiable but not necessarily invertible. In [4], a diffeomorphism is represented using a *local kernel* to pull back the Riemannian metric from one manifold onto the other. With respect to the intrinsic geometry of the local kernel, the manifolds are isometric, and the isometry can be represented by a linear map between the eigenfunctions of the respective Laplacian operators. In this paper, we consider the more difficult case when the manifolds are not diffeomorphic, so that one manifold may even be higher dimensional than the other. This is typically the case with feature maps, since the data space may contain irrelevant variables. This implies that the data manifold dimension, d, may be greater than the feature manifold dimension, $d_{\mathcal{N}}$. In the annulus example the data space is two dimensional and both the feature (radius, r) and the irrelevant variable (angle, θ) are one dimensional.

The challenge of having irrelevant variables is that it violates the fundamental assumption of differential geometry, namely that it is local. This is because data points which differ only in the irrelevant variables will be far away in the data space and yet have the same feature values. This fundamental issue is independent of the amount of data available and is illustrated in Fig. 1. Namely, if the feature of interest is the radius of an annulus, then points on opposite sides of the annulus are closely related with respect to this feature of interest. Conversely, points which are far away in the feature space may appear relatively close in data space;

Please cite this article in press as: T. Berry, J. Harlim, Iterated diffusion maps for feature identification, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.08.005