Accepted Manuscript

Parsimonious representation of nonlinear dynamical systems through manifold learning: A chemotaxis case study

Carmeline J. Dsilva, Ronen Talmon, Ronald R. Coifman, Ioannis G. Kevrekidis

PII: S1063-5203(15)00094-9
DOI: http://dx.doi.org/10.1016/j.acha.2015.06.008
Reference: YACHA 1054

To appear in: *Applied and Computational Harmonic Analysis*

Received date: 26 May 2015
Accepted date: 19 June 2015

Please cite this article in press as: C.J. Dsilva et al., Parsimonious representation of nonlinear dynamical systems through manifold learning: A chemotaxis case study, *Appl. Comput. Harmon. Anal.* (2015), http://dx.doi.org/10.1016/j.acha.2015.06.008

Parsimonious Representation of Nonlinear Dynamical Systems Through Manifold Learning: A Chemotaxis Case Study

Carmeline J. Dsilva[a], Ronen Talmon[b], Ronald R. Coifman[c], Ioannis G. Kevrekidis[a,d,*]

[a]*Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ, 08540, USA*
[b]*Department of Electrical Engineering, Technion - Israel Institute of Technology, Haifa, 3200003, Israel*
[c]*Department of Mathematics, Yale University, New Haven, CT, 06520, USA*
[d]*Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ, 08540, USA*

Abstract

Nonlinear manifold learning algorithms, such as diffusion maps, have been fruitfully applied in recent years to the analysis of large and complex data sets. However, such algorithms still encounter challenges when faced with real data. One such challenge is the existence of "repeated eigendirections," which obscures the detection of the true dimensionality of the underlying manifold and arises when several embedding coordinates parametrize the same direction in the intrinsic geometry of the data set. We propose an algorithm, based on local linear regression, to automatically detect coordinates corresponding to repeated eigendirections. We construct a more parsimonious embedding using only the eigenvectors corresponding to unique eigendirections, and we show that this reduced diffusion maps embedding induces a metric which is equivalent to the standard diffusion distance. We first demonstrate the utility and flexibility of our approach on synthetic data sets. We then apply our algorithm to data collected from a stochastic model of cellular chemotaxis, where our approach for factoring out repeated eigendirections allows us to detect changes in dynamical behavior and the underlying intrinsic system dimensionality directly from data.

Keywords: diffusion maps, repeated eigendirections, chemotaxis

1. Introduction

In recent years, data mining algorithms have proven useful for many disciplines and applications [1, 2, 3, 4, 5, 6, 7, 8]. For multiscale dynamical systems in particular, data-driven methodologies are essential for assisting in model reduction when simple macroscopic models cannot be analytically obtained due to the system complexity [9, 10, 11, 12]. For simulations or experimental observations of such systems, the detailed, microscale evolving system state is very high-dimensional, and the reduction to useful macroscale dynamics is not obvious. In such cases, data obtained from observations and/or simulations of the dynamical system combined with data mining methodologies can lead to a low-dimensional description which not only provides insight into the underlying dynamics, but also serves as a first step in constructing macroscale models consistent with the observed microscale behavior.

Several manifold learning algorithms obtain parametrizations of the data through the spectral analysis of a Laplace operator [13, 14, 15, 16]. The data is then embedded in a new low-dimensional coordinate system given by the eigenvectors of this operator. The premise is that these coordinates, obtained in a data-driven manner, are the right macroscopic "observables", i.e., the variables which parametrize the macroscale dynamical behavior of the system, thus enabling the construction of a reduced macroscale model.

Such manifold learning algorithms were initially applied to synthetic data sets, to illustrate their geometric properties and flexibility [14, 17]. More recently they have been applied to experimental as well as

*Corresponding author
Email address: yannis@princeton.edu (Ioannis G. Kevrekidis)