Computers and Mathematics with Applications 65 (2013) 1535–1557

Contents lists available at SciVerse ScienceDirect

Computers and Mathematics with Applications

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

Noisy dynamic simulations in the presence of symmetry: Data alignment and model reduction

Benjamin Sonday [a], Amit Singer [a,b], Ioannis G. Kevrekidis [a,c,*]

[a] *Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA*
[b] *Department of Mathematics, Princeton University, Princeton, NJ 08544, USA*
[c] *Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, USA*

ARTICLE INFO

Keywords:
Dimensionality reduction
Heat kernel
Local principal component analysis
Alignment

ABSTRACT

We process snapshots of trajectories of evolution equations with intrinsic symmetries, and demonstrate the use of recently developed eigenvector-based techniques to successfully quotient out the degrees of freedom associated with the symmetries in the presence of noise. Our illustrative examples include a one-dimensional evolutionary partial differential (the Kuramoto–Sivashinsky) equation with periodic boundary conditions, as well as a stochastic simulation of nematic liquid crystals which can be effectively modeled through a nonlinear Smoluchowski equation on the surface of a sphere. This is a useful first step towards data mining the symmetry-adjusted ensemble of snapshots in search of an accurate low-dimensional parametrization and the associated reduction of the original dynamical system. We also demonstrate a technique (Vector Diffusion Maps) that combines, in a single formulation, the symmetry removal step and the dimensionality reduction step.

1. Introduction

High-dimensional dynamical systems are often characterized by low-dimensional long-term dynamic behavior. Obtaining reduced-dimensionality models consistent with this behavior is clearly useful both in analysis and in computations. While such model reduction can be based on properties of the dynamics (e.g. Center-Manifold or Lyapunov–Schmidt reduction, see [1,2], or Inertial and Approximate Inertial Manifolds, see [3–9]), semi-empirical methods based on data-mining are also enjoying extensive use in applications (e.g. PCA/POD-Galerkin methods, see [10–13]). As nonlinear extensions of Principal Component Analysis are developed (e.g. techniques like Isomap, Local Linear Embedding, Laplacian Eigenmaps/Diffusion Maps, etc., see [14–18]), the necessity of linking these nonlinear data reduction techniques with dynamic model reduction naturally arises.

When the data set of interest consists of snapshots of trajectories of dynamical systems with symmetry, factoring out this symmetry is an established first step (in theory, in computations, as well as in PCA-based data mining); the use of the so-called *template functions* in this context has been described by Rowley and coworkers (e.g. [19,20], see also [21,22]). In this paper we explore the application of recently developed computational approaches to symmetry removal (factoring out symmetry, *alignment* of the data) for (noisy) high-dimensional dynamical system data. Our illustrative examples include (1) the discretization of a well-known spatiotemporal pattern-forming partial differential equation (PDE), the Kuramoto–Sivashinsky Equation (KSE) in one spatial dimensional and with periodic boundary conditions (with associated symmetry group SO(2)); and (2) a stochastic simulation of a nonlinear 2D Smoluchowski equation, where the evolution of

* Corresponding author at: Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA. Tel.: +1 609 532 0772.
E-mail address: yannis@princeton.edu (I.G. Kevrekidis).

0898-1221/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2013.01.024

the orientational distribution function of an ensemble of nematic liquid crystals is modeled on the sphere (with associated symmetry group SO(3)). In both cases noise is present in the data; in the KSE case the noise is added externally (by us); in the nematic liquid crystal case the noise comes from the stochastic simulation of a finite ensemble of representative particles.

The essential step in factoring out the relevant symmetries involves relating each snapshot in the data to each other snapshot (in effect, using each snapshot as the "alignment template" for every other snapshot); using these pairwise relations to perform a global alignment can be formulated as an optimization problem that is fruitfully relaxed to an eigenproblem (hence the term "eigenvector method", see [23–25]).

In one of our examples (the KSE) we will also demonstrate the combination of this alignment with a second, data mining (dimensionality reduction) step; the combination carries the name of *Vector Diffusion Maps* [26] and has potential advantages over the two step approach (first alignment and then reduction). The data set corresponding to the snapshots of the dynamical system is usually modeled as lying on a low dimensional manifold $\mathcal{M}$. In the presence of a symmetry group G (such as SO(2) or SO(3)), vector diffusion maps provide a natural framework to organize the data in the quotient space $\mathcal{M}/G$. The affinities between data points are related to their correlation when they are optimally aligned, and the information about the optimal alignment transformation (the group element) is also encoded in this framework. The advantage of working in the quotient space $\mathcal{M}/G$ stems from its lower dimensionality compared to the original manifold $\mathcal{M}$, giving rise to improved dimensionality reduction, noise robustness, and the need for less data.

The paper is organized as follows. In Section 2, we give an overview of the alignment problem and briefly review template-based methods. Next, Section 3 summarizes the eigenvector method and some of its relevant mathematical properties. Sections 4 and 5 are devoted to applying and comparing template-based approaches and the eigenvector method to our two prototypical examples. Finally, in Sections 6 and 7, we demonstrate the use of two dimensionality reduction techniques, Diffusion Maps and Vector Diffusion Maps, on the modulated traveling wave data of Section 5.

2. Description of the problem

For physical systems possessing symmetry, there may be several equivalent realizations of what is effectively the same system state (whether a steady/stationary state or an time-instance or *snapshot* during a transient simulation); these realizations are related by some underlying symmetry group. When such systems with symmetry evolve in time, their dynamics are equivariant with respect to the appropriate symmetry group. Consider a function $u(\theta, t)$ on the unit circle evolving according to some spatially invariant differential operator $\mathcal{D}$ via an equation of the form

$$u_t = \mathcal{D}(u). \tag{1}$$

This equation is equivariant in the sense that

$$\mathcal{D}(S_c[u]) = S_c[\mathcal{D}(u)], \tag{2}$$

where $S_c[v](\theta) = v(\theta + c)$ is the *shift operator* on spatially periodic functions; starting at a particular snapshot, evolving the dynamics for some time and shifting the final state by c is the same as the result of shifting the initial snapshot by c and then evolving the dynamics from the shifted initial condition (in other words, the differential operator $\mathcal{D}$ commutes with the shift operator S_c).

Suppose we take M snapshots of u at M different times, $\{u(\theta, t_k)\}_{k=1}^{M}$. If $u(\theta, t)$ is not changing its shape, but simply traveling around the unit circle (for example, when $\mathcal{D}(u) = \omega u_\theta$), we may identify each snapshot with some angle $\theta \in [0, 2\pi)$. By rotating each of these snapshots back by the angle θ with which it has been identified, we obtain a set of *identical* system snapshots (thereby removing one degree of freedom from the evolving system).

The removal of this degree of freedom allows us to perform certain tasks, such as denoising a collection of snapshots through averaging more easily (in [24,25], a similar procedure is used on cryo-EM data). In the case where $u(\theta, t)$ is evolving its shape in addition to traveling (for example, when $\mathcal{D}(u) = \omega u_\theta + \mathcal{E}(u)$, where $\mathcal{E}(u)$ is some other nonlinear spatially invariant operator), removing this traveling degree of freedom from the simulation can significantly assist our understanding of the dynamics. Fig. 1 illustrates this simplification. For instance, when one uses *diffusion maps* to explore whether the simulation data are intrinsically low-dimensional, and to find good "coarse" parametrizations for them (see, e.g., [27–31]), removing the symmetry results in a more parsimonious description of the dynamics (an embedding in a lower-dimensional space), which may also be successfully deduced with far less data.

Now suppose we have an ensemble of M snapshots, but we do not know the members of the underlying symmetry group with which each snapshot is to be identified. We wish to perform this association of snapshots with symmetry group elements; in other words we wish to globally align the M snapshots. (Here the colloquial expression *alignment* comes from the simple conceptual example of rotationally invariant functions on the unit circle; the possible rotation angles can be "strung" along a line between 0 and 2π.)

Normally, this global alignment (the computation of the symmetry group element identified with each snapshot) may be accomplished numerically through the use of a well-chosen *template function* (see Fig. 1 and, e.g., [19,20]). For instance, in our running example of snapshots $\{u(\theta, t_k)\}_{k=1}^{M}$, one finds the alignments $\{\theta_k\}_{k=1}^{M}$ which align each snapshot with a template $T(\theta)$ by simply setting

$$\theta_k = \underset{c}{\operatorname{argmin}} \, \|T(\theta) - u(\theta + c, t_k)\|^2. \tag{3}$$