Journal of Computational Physics 372 (2018) 281–298

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

# Data-driven closures for stochastic dynamical systems

Catherine Brennan, Daniele Venturi *

*Department of Applied Mathematics and Statistics, University of California Santa Cruz, Santa Cruz, CA 95064, United States of America*

ARTICLE INFO

*Article history:*
Received 6 April 2018
Received in revised form 10 June 2018
Accepted 11 June 2018
Available online xxxx

*Keywords:*
Reduced-order stochastic modeling
Mori–Zwanzig formulation
Nonlinear dynamics
Liouville equation

ABSTRACT

We develop a new data-driven closure approximation method to compute the statistical properties of quantities of interest in high-dimensional stochastic dynamical systems. The proposed framework relies on estimating system-dependent conditional expectations from sample paths or experimental data, and then using such estimates to compute data-driven solutions to exact probability density function (PDF) equations. We also address the important question of whether enough useful data is being injected into the exact PDF equation for the purpose of computing an accurate numerical solution. Numerical examples are presented and discussed for prototype nonlinear dynamical systems and models of systems biology evolving from random initial states.



## 1. Introduction

High-dimensional stochastic dynamical systems arise naturally in many areas of engineering, physical sciences and mathematics. Whether it is a physical system being studied in a lab or an equation being solved on a computer, the full state of the system is often intractable to handle in all its complexity. Instead, it is often desirable to reduce such complexity by moving from a full model of the dynamics to a reduced-order model that involves only a small number of quantities of interest. Such quantities of interest may represent specific features of the system, e.g., the sensitivity of tumor populations to chemo-treatment in stochastic models tumoral cell growth [1,11], or the viscous dissipation in intertial range of fully developed turbulence [21,29]. The dynamics of the quantities of interest may be simpler than that of the entire system, although the underlying law by which they evolve in space and time is often quite complex. Nevertheless, approximation of such law can in many cases allow us to avoid performing simulation of the full system and solve directly for the quantities of interest. In this paper, we aim at providing a new general framework to compute the probability density function (PDF) of such quantities of interest based on data-driven closure approximations. To introduce the methodology, consider the following $N$-dimensional nonlinear dynamical system evolving on a smooth manifold $\mathcal{M} \subseteq \mathbb{R}^N$

$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt} = \boldsymbol{G}(\boldsymbol{x}) \\ \boldsymbol{x}(0) = \boldsymbol{x}_0 \end{cases}, \tag{1}$$

where $\boldsymbol{x}_0$ is a random initial state with known probability density function $p(\boldsymbol{x}_0)$. Non-autonomous systems driven by finite-dimensional (time-dependent) random noise can be always written in the form (1), by augmenting the number of phase variables (see, e.g., [32,2]). Suppose we are interested in the dynamics of a real-valued phase space function

* Corresponding author.
*E-mail address:* venturi@ucsc.edu (D. Venturi).

https://doi.org/10.1016/j.jcp.2018.06.038
0021-9991/

$$u(\boldsymbol{x}) = \mathcal{M} \to \mathbb{R} \qquad \text{(quantity of interest).} \tag{2}$$

The exact dynamics of the quantity of interest (2) can be expressed as

$$u(\boldsymbol{x}(t, \boldsymbol{x}_0)) = \exp\left[t\mathcal{K}(\boldsymbol{x}_0)\right] u(\boldsymbol{x}_0), \qquad \mathcal{K}(\boldsymbol{x}_0) = \sum_{k=1}^{N} G_k(\boldsymbol{x}_0)\frac{\partial}{\partial x_{0k}}, \tag{3}$$

where $\boldsymbol{x}(t, \boldsymbol{x}_0)$ denotes the flow map [37] generated by the system (1), and $\exp[t\mathcal{K}(\boldsymbol{x}_0)]$ is the Koopman operator [16,9,18]. Differentiation of (3) with respect to time yields

$$\frac{\partial u(t, \boldsymbol{x}_0)}{\partial t} = \boldsymbol{G}(\boldsymbol{x}_0) \cdot \nabla u(t, \boldsymbol{x}_0), \qquad u(0, \boldsymbol{x}_0) = u(\boldsymbol{x}_0), \tag{4}$$

where the gradient is with respect to the variables $\boldsymbol{x}_0$. The solution to the initial value problem (4) allows to determine the dynamics of the quantity of interest corresponding to any initial condition $\boldsymbol{x}_0$.

**Example 1.1.** Setting $u(\boldsymbol{x}(t, \boldsymbol{x}_0)) = x_i(t, \boldsymbol{x}_0)$ for $i = 1, \ldots, N$ yields

$$\frac{\partial \boldsymbol{x}(t, \boldsymbol{x}_0)}{\partial t} = \boldsymbol{G}(\boldsymbol{x}_0) \cdot \nabla \boldsymbol{x}(t, \boldsymbol{x}_0), \qquad x(0, \boldsymbol{x}_0) = \boldsymbol{x}_0. \tag{5}$$

This system of linear PDEs, together with the initial condition $\boldsymbol{x}(0, \boldsymbol{x}_0) = \boldsymbol{x}_0$, allows us to compute the flow map generated by (1).

The dual of the Koopman operator $\exp(t\mathcal{K})$ with respect to the inner product

$$\langle f, g \rangle = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\boldsymbol{x}_0) g(\boldsymbol{x}_0) p(\boldsymbol{x}_0) d\boldsymbol{x}_0 \tag{6}$$

can be written in the form $\exp(t\mathcal{L})$, where

$$\mathcal{L}(\boldsymbol{x})\phi = -\nabla \cdot (\boldsymbol{G}(\boldsymbol{x})\phi(\boldsymbol{x})). \tag{7}$$

The operator semigroup $\exp(t\mathcal{L})$ is known as Frobenious–Perron (or transfer) operator. It characterizes the evolution of the joint PDF of the solution to (1), i.e.,

$$p(\boldsymbol{x}, t) = e^{t\mathcal{L}(\boldsymbol{x})} p(\boldsymbol{x}, 0). \tag{8}$$

Differentiation of (8) with respect to time yields the well-known Liouville transport equation [27,32,33]

$$\frac{\partial p(\boldsymbol{x}, t)}{\partial t} + \nabla \cdot (\boldsymbol{G}(\boldsymbol{x}) p(\boldsymbol{x}, t)) = 0. \tag{9}$$

Computing the numerical solution of (9) can be quite challenging due to complications with high-dimensionality, multiple scales, and conservation properties [20]. In fact, from a mathematical viewpoint, (9) is a hyperbolic conservation law in as many variables as the dimension of the system (1).

**Remark 1.1.** By using the method of characteristics [26], it is straightforward to obtain the following formal solution to (9)

$$p(\boldsymbol{x}, t) = p_0\left(\boldsymbol{x}_0(\boldsymbol{x}, t)\right) \exp\left(-\int_0^t \nabla \cdot \boldsymbol{G}\left(\boldsymbol{x}(\tau, \boldsymbol{x}_0)\right) d\tau\right). \tag{10}$$

Here, $p_0(\boldsymbol{x}) = p(\boldsymbol{x}, 0)$, while $\boldsymbol{x}_0(\boldsymbol{x}, t)$ denotes the inverse flow map generated by (1). Note that this expression provides a representation of the Frobenious–Perron operator semigroup (8).

This paper is organized as follows. In Section 2 we develop reduced-order PDF equations for arbitrary quantities of interest (2) and discuss their mathematical properties. In particular, we comment extensively on the closure problem arising from the dimension reduction procedure and relate it with the need of computing/estimating conditional expectations. In Section 3 we propose a robust procedure to compute such conditional expectations based on sample paths of (1), or experimental data. This opens the possibility to compute data-driven solutions to reduced-order PDF equations, and also estimate the memory integral arising in the Mori–Zwanzig formulation [31,38,39] (Section 5). In Section 4 we develop a new paradigm to measure the information content of data. Such paradigm allows us to infer, in particular, whether we have enough data to accurately close the reduced-order PDF equation for the quantity of interest. Finally, in Section 6 we demonstrate the proposed data-driven closure approximation method in applications to a high-dimensional nonlinear dynamical system and a drug resistant malaria propagation model.