Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2018-0174

Vol. **25**, No. 5, pp. 1259-1301
May 2019

Physics-Constrained, Data-Driven Discovery of Coarse-Grained Dynamics

Lukas Felsberger[1] and Phaedon-Stelios Koutsourelakis[1,*]

[1] *Continuum Mechanics Group, Department of Mechanical Engineering, Technical University of Munich, Boltzmannstrasse 15 85748, Garching, Germany.*

Received 12 February 2018; Accepted (in revised version) 6 July 2018

Abstract. The combination of high-dimensionality and disparity of time scales encountered in many problems in computational physics has motivated the development of coarse-grained (CG) models. In this paper, we advocate the paradigm of data-driven discovery for extracting governing equations by employing fine-scale simulation data. In particular, we cast the coarse-graining process under a probabilistic state-space model where the transition law dictates the evolution of the CG state variables and the emission law the coarse-to-fine map. The directed probabilistic graphical model implied, suggests that given values for the fine-grained (FG) variables, probabilistic inference tools must be employed to identify the corresponding values for the CG states and to that end, we employ Stochastic Variational Inference. We advocate a sparse Bayesian learning perspective which avoids overfitting and reveals the most salient features in the CG evolution law. The formulation adopted enables the quantification of a crucial, and often neglected, component in the CG process, i.e. the predictive uncertainty due to information loss. Furthermore, it is capable of reconstructing the evolution of the full, fine-scale system. We demonstrate the efficacy of the proposed framework in high-dimensional systems of random walkers.

AMS subject classifications: 62F15, 82C21, 82C80

Key words: Coarse-graining, dynamics, Bayesian, non-equilibrium, data-driven.

1 Introduction

The present paper is concerned with the discovery of data-driven, dynamic, stochastic coarse-grained models from fine-scale simulations with a view of advancing multiscale modeling. Many problems in science and engineering are modeled by high-dimensional systems of deterministic or stochastic, (non)linear, microscopic evolution laws (e.g. ODEs). Several such examples are encountered in quantum mechanical models [1], in the atomistic simulation of materials [2], in complex flows [3–5], and in agent-based models [6].

*Corresponding author. *Email addresses:* p.s.koutsourelakis@tum.de (P. S. Koutsourelakis), lukas.felsberger@cern.ch (L. Felsberger)

Their solution is generally dominated by the smaller time scales involved even though the outputs of interest might pertain to time scales that are greater by several orders of magnitude [7]. The combination of high-dimensionality and disparity of time scales has motivated the development of coarse-grained formulations. These aim at constructing a much lower-dimensional model that is practical to integrate in time and can adequately predict the outputs of interest over the time scales of interest. The literature on this topic is vast and many categorizations are possible. This paper focuses on *data-driven* strategies [8–18] where (short) simulations of the original, fine-grained (FG) or full-order system of equations are used in order to learn (or infer) the right coarse-grained (CG) model. This is consistent with the emergence of data-driven discovery, commonly referred to as the fourth paradigm in science [19]. Extracting governing equations from data is a central challenge in a wide variety of physical and engineering sciences as in climate science, neuroscience, ecology, finance, and epidemiology, where models (or closures) often remain elusive [20–23].

The challenge in multiscale physical systems such as those encountered in non-equilibrium statistical mechanics [24], is even greater as apart from the identification of an effective model, it is crucial to discover a good set of CG state variables [25]. The latter problem has been traditionally addressed separately from the construction of the CG model, using dimensionality reduction techniques, (e.g. [26]) but in this work we offer a holistic treatment.

The starting point of all (to the best of our knowledge) CG schemes is the specification of a set of coarse (or reduced) state variables with respect to which a CG model is prescribed/found. These variables are selected on the basis of the analysis objectives, or because they are known to be "slow" either through physical insight or using rigorous statistical learning tools (e.g. dimensionality reduction). The specification of these coarse variables is generally done through a projection (or restriction), many-to-one operator from the fine-scale variables to the coarse. For problems exhibiting time-scale separation and when the CG state variables identified correspond to slowly-evolving features, it has been established that Markovian CG models are suitable and several numerical strategies have proven successful [27, 28]. In the absence of scale separation however, memory effects (dissipation) and thermal noise play an important role. The Mori-Zwanzig (MZ) formalism which was originally developed in the context of irreversible statistical mechanics [29, 30] but has been extended to general systems of ODEs (e.g. [31–34]) provides a rigorous mathematical foundation. Setting aside the largely unsolved computational challenges associated with finding the terms in the Generalized Langevin Equation (GLE) [35, 36] prescribed by MZ, it is in principle a perfect scheme, i.e. it is capable of capturing exactly the dynamics of the state variables that are resolved and with respect to which a closed system of equations is written. The solution of non-Markovian system of equations (due to the memory term) poses itself several difficulties as it necessitates storing the history of the CG variables. Interestingly, in many instances, this non-Markovian system can be mapped onto an, albeit higher-dimensional, Markovian, system with additional degrees of freedom corresponding to auxiliary or extended state