Computational Materials Science 152 (2018) 125–133

Contents lists available at ScienceDirect

Computational Materials Science

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

Fast predictive models based on multi-fidelity sampling of properties in molecular dynamics simulations

M. Razi[a,*], A. Narayan[a], R.M. Kirby[a], D. Bedrov[b]

[a] Scientific Computing and Imaging Institute, University of Utah, United States
[b] Department of Materials Science and Engineering, University of Utah, United States

ARTICLE INFO

Keywords:
Model-order reduction
Molecular dynamics
Multi-fidelity models
Parameter estimation
Surrogate modeling

ABSTRACT

In this paper we introduce a novel approach for enhancing the sampling convergence for properties predicted by molecular dynamics. The proposed approach is based upon the construction of a multi-fidelity surrogate model using computational models with different levels of accuracy. While low fidelity models produce result with a lower level of accuracy and computational cost, in this framework they can provide the basis for identification of the optimal sparse sampling pattern for high fidelity models to construct an accurate surrogate model. Such an approach can provide a significant computational saving for the estimation of the quantities of interest for the underlying physical/engineering systems. In the present work, this methodology is demonstrated for molecular dynamics simulations of a Lennard-Jones fluid. Levels of multi-fidelity are defined based upon the integration time step employed in the simulation. The proposed approach is applied to two different canonical problems including (i) single component fluid and (ii) binary glass-forming mixture. The results show about 70% computational saving for the estimation of averaged properties of the systems such as total energy, self diffusion coefficient, radial distribution function and mean squared displacements with a reasonable accuracy.

1. Introduction

Accurate sampling of the evolution of system of interest in molecular dynamics (MD) simulations can be very challenging. Thus, despite their outstanding predictive power and application in different areas of science and engineering, MD simulations can be carried out over a limited timescale. Since there is a demand for running these simulations for much longer timescales, particularly for studying systems with rough energy landscape and long relaxation processes, several approaches have been proposed in the literature to address this problem.

One of the traditional ways of dealing with this issue is to apply transition-state-theory, compute the rate relevant for infrequent events, and eventually obtain an estimation of the long-time MD results [1]. However, prior determination of all important reaction paths is very difficult. Hence, researchers have proposed different modifications of the potential energy models for MD simulations to overcome the energy barriers much faster. In these series of approaches, through raising the potential energy surfaces in the regions of potential minima, in which standard MD simulations spend a large portion of their computational time, faster exploration of potential energy landscape becomes possible. Hence the simulation moves faster over potential barriers and is able to capture infrequent-event transitions and eventually the equilibrium status of the system in much less computational time [2,3]. This group of approaches, which are characterized by interatomic energy manipulation, are known as the accelerated molecular dynamics (AMD) method. Parallel-replica dynamics method using process parallelization, hyperdynamics approach based upon importance sampling, and temperature-accelerated dynamics by adaptive assignment temperature to transitions are among the well known AMD approaches developed [3]. Similar to the hyperdynamics method, other authors have explored the alteration of potential landscape by proposing a bias potential function as a means for accelerating the MD simulation [2,4]. However, implementation of these methods requires an in-depth understanding of underlying molecular dynamics processes and often performing a series of rigorous computational procedures.

Recently, application of predictive algorithms in the area of molecular dynamics has proven to achieve accurate and computationally efficient long timescale analysis. These approaches span from functional uncertainty quantification (UQ) [5] to multi-fidelity machine learning models [6]. In this context, Reeve and Strachan introduced a novel technique to apply functional UQ to Lennard-Jones two-body interaction model for prediction of high order interatomic interactions [5]. In spite of using a perturbative technique for the numerical evaluation of function derivatives, this approach is computationally

* Corresponding author at: Scientific Computing and Imaging Institute, University of Utah, 72 Central Campus Dr, RM: 2654, Salt Lake City, UT 84112, United States.
E-mail addresses: manirazi@sci.utah.edu (M. Razi), akil@sci.utah.edu (A. Narayan), kirby@sci.utah.edu (R.M. Kirby), d.bedrov@utah.edu (D. Bedrov).

https://doi.org/10.1016/j.commatsci.2018.05.029
Received 28 February 2018; Received in revised form 12 May 2018; Accepted 17 May 2018

demanding since a sufficiently large number of samples from low-fidelity models is required. This is mainly to ensure that the low-fidelity model explores enough of phase space of the high-fidelity model. Furthermore, this approach works only when the discrepancies between high-fidelity and low-fidelity potential energy models stay within reasonable bounds.

In another recent research study, multi-fidelity machine learning regression models have been used for the accurate bandgap prediction of solid materials [6]. This approach, which is based upon a Gaussian process (GP) regression framework, involves the application of co-kriging statistical learning on several bandgap predictive models with different levels of fidelity. This requires having sufficient independent data for parameter estimation and the assumption that the Gaussian process properly describes the underlying molecular dynamics. Hence, any deviation of underlying physics and limited availability of data can lead to a poor parameter estimation.

In the present paper, considering the challenges in computational complexity of predictive models and limited availability of even low-fidelity data in practice, a stochastic collocation methodology with multi-fidelity [7,8] is applied to MD simulation in order to built accurate surrogate models. This approach does not require any *a priori* assumption about the probability measure of the underlying physics that the discrepancies between low-fidelity and high-fidelity model can be much larger than what can be recovered by standard predictive surrogate modeling tools such as spectral (polynomial based) or Bayesian surrogates. The examples of such a large discrepancies between the mathematical models of low- and high-fidelity and the effectiveness of similar multi-fidelity surrogate model constructions in other areas including frequency-modulated trigonometric functions [7], heat driven cavity flows [9] and irradiated particle-laden turbulence [10] have been discussed in the literature. This surrogate modeling approach is also designed to work with a limited number of samples and provide an optimal sampling strategy for the high-fidelity MD simulations. Moreover, compared with the commonly used approaches for acceleration of MD simulations, the proposed method works well not only with small data availability but also with significant difference between high- and low-fidelity models. As long as the low-fidelity model reflects the impact of the variation of model parameters in parameter space (not solution space), it can be used in this framework. In order to demonstrate the proposed multi-fidelity predictive method, it is applied to two canonical problems involving output parameter estimation of MD simulation for (i) one- and (ii) two-component systems. In this study, the integration time step of the MD simulation defines its level of fidelity. Here, by using the proposed multi-fidelity surrogate modeling approach, the goal is to accelerate parameter exploration for a given MD simulation setup. In this light, the resultant acceleration of the MD simulation does directly benefit situations when the MD simulation has a longer terminal time and hence provides an effective tool for faster exploration of time-scales in the defined phase space. As such, this approach can be considered as an efficient alternative to AMD methods, particularly when obtaining a series of MD solutions for different sets of parameters (and not only one) is desirable. The results of our experiments with the application of the proposed approach to both test problems indicate both accuracy and computational efficiency of constructed data driven predictive models. The aforementioned canonical problems are considered for the sole purpose of proof of concept and demonstration of the proposed approach, which is used in the area of molecular dynamics simulations for the first time. In this sense, there is no limitation to apply this approach to more complex MD problems and the authors plan to investigate those cases in their future works. It is also worth noting that the processes of uniform random sampling for the low-fidelity model and important sampling for the high-fidelity model in the phase space lend itself nicely to parallel implementation. By taking advantage of this feature, one can expect some additional acceleration in computational speed for the construction of this multi-fidelity predictive model.

The present manuscript is organized as follows. In Section 2, a brief description of selected model system is provided. This follows by a detailed theoretical foundation of the proposed multi-fidelity in the subsequent section. Next, the results and discussion for two canonical problems are presented in Section 4. Finally, concluding remarks are provided as the final section of this paper.

2. Theoretical foundation

2.1. Molecular dynamics simulation model

The classical computational algorithm for performing molecular dynamics simulations has been developed based upon the Lagrangian methodology of tracking particle dynamics based on Newton's equations of motion. Numerical integration of these equations provides an accurate evaluation of the time evolution of a molecular system and hence the system's quantities of interest. While there are many integration schemes available in the literature, the velocity Verlet algorithm [11] is used in the present work due to its popularity, reasonable accuracy and simplicity of implementation. The Lennard-Jones (LJ) potential [11] was used for simulating the interatomic interactions in the MD system:

$$u(r_{ij}) = 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right], \tag{1}$$

where r_{ij} denotes the pairwise distance between particles i and j, ϵ is the potential well depth and σ defines the length scale for this pairwise interatomic interaction model. For two-species interactions this length scale becomes $\sigma = \sqrt{\sigma_1\sigma_2}$. Here, in the second test problem, we consider a two-component MD system with $\frac{\sigma 1}{\sigma_2} = 0.9$ and same potential well depth.

In the process of molecular dynamic simulation using any interatomic interaction model, when the size of the integration time-step is large, the major challenge is going to be the stability of the MD integration scheme. If the time-step is too large, then a one time step integration may predict significant overlap between molecules/atoms which will lead to huge (unphysical) repulsive forces and large displacements on the next time step, which in turn will lead to even larger overlap and more unphysical forces on the next time step. To prevent such divergence of integration scheme, we capped the magnitude of repulsive interactions for closely approaching atoms. For the Lennard-Jones potential, such capping can be implemented straightforwardly by modifying the potential at short distance. As such, for both test cases in this work, a potential energy cap is considered when the ratio of $\frac{\sigma}{r_{ij}}$ exceeds 1.2. For more complex systems which may involve electrostatic interactions, the capping should include modifications of charge-charge, charge-induced dipole, and induced dipole-induced dipole interactions. These modifications would require a slightly more coding and data management, but will be similar to current schemes of excluding or scaling of intramolecular electrostatic interactions between atoms connected by bonds, bends, and dihedrals.

2.2. Bi-fidelity model construction

In the context of MD simulations, the size of time-step for the integration procedure plays a crucial rule in the computational cost of achieving large time-scale predictions. The drawback of choosing a large time-step to accelerate the MD computations is the loss of accuracy and more often numerical stability of the solution. Here, our goal is to build a multi-fidelity surrogate model, for which the levels of fidelity are defined based upon the time-step size of the corresponding MD simulation. The first step in building such a model is to define a region of parameter space where one needs to estimate state variables with a reasonable accuracy. Choosing a large-time step and running the low-