Studies of a monatomic Lennard-Jones system at slow cooling by molecular dynamics simulations and regression analysis

A R Farkhutdinov

1Kazan Federal University, Kazan, Russia

farhutdinovalbert@yandex.ru

Abstract. The procedure of radial distribution function calculation at low cooling rates without a substantial increase in the computational cost is proposed in this paper. Results of its application to monatomic Lennard-Jones system at cooling rates \( \gamma = 4 \times 10^{-4} \ldots 2 \times 10^{-2} \varepsilon/k_B \tau \), density \( \rho \approx 0.75 \sigma^{-3} \) and temperature range \( T = 0.3 \ldots 1.0 \varepsilon/k_B \) using various regression models are given. It is shown that \( k \)-nearest neighbors regression yields the minimum deviation from the results of molecular dynamics simulations in comparison with other models.

1. Introduction

Molecular dynamics simulations allow obtaining comprehensive information about condensed matters. These methods complement the results of traditional experiments in spectroscopy, diffraction, microscopy, etc. On the other hand, they provide a reliable basis for the development of theoretical concepts [1–8].

However, some phenomena computation is challenging due to a huge gap in time scales and cooling rates between experiments and conventional molecular dynamics simulations. Thus, the development of methods allowing simulating a many-body system at realistic conditions is relevant. This work is concerned with predicting Lennard-Jones fluid structural properties, in particular, radial distribution function (RDF). First attempts to model RDF of the Lennard-Jones fluid have been based on fitting empiric expressions to simulation data. Goldsman equation [9] accurately predicts RDF in certain limited conditions however its usage is mathematically complicated. The expression by Morsali et al. [10] involves fewer parameters however its accuracy descends with density decrease. Bamdad et al. [11] introduced two different expressions for the first peak of RDF and the rest. Lack of continuity is also a problem in the expressions proposed by Matteoli and Mansoori [12].

Another approach to predict the RDF of simple fluids is to use integral equation methods [13–16]. However, they have not produced acceptable results for the Lennard-Jones fluid. This is because the functional form of the integral equation must be modified specifically to treat the Lennard-Jones system.

Well-Tempered metadynamics (WTmetaD) approach to compute nucleation rates in the Lennard-Jones system was proposed by Salvalaglio et al [17]. WTmetaD is an enhanced sampling method conventionally used to compute free energy surfaces in a variety of contexts. According to this work, the timescale limitations that plague the simulation of nucleation phenomena can be overcome using WTmetaD simulations. In the case of argon condensation, this implies being able to reach timescales of the order of hours with ordinary computational resources.
Machine learning (ML) approaches have been applied to examine structural correlations in a Lennard-Jones fluid. Moradzadeh and Aluru [18] have trained a neural network to predict what values of the Lennard-Jones parameters will give a specific form of the RDF. They have also trained an autoencoder network to determine the RDF of simple fluids from a limited number of atomistic configurations. Craven et al [19] have previously developed a ML method based on segmented linear regression and multivariate function decomposition that can predict the RDF of simple fluids with significantly increased accuracy in comparison to traditional theoretical approaches.

In this work, we propose the RDF calculation algorithm for the monatomic Lennard-Jones system using various regression models. It is shown below that k-nearest neighbors regression yields the minimum deviation from the results of molecular dynamics simulations in comparison to other models.

The paper is structured as follows; at first, the details of the studied system and simulation parameters are reported. Then regression procedure is described. Finally, results are discussed.

2. Model and simulation

A monatomic system of $N = 1372$ atoms in three dimensions is considered. Particles interact with each other isotropically via the truncated and shifted Lennard-Jones potential [20]:

$$U(r_{ij}) = \begin{cases} U_{LJ}(r_{ij}) - U_{LJ}(r_C), & r_{ij} < r_C \\ 0, & r_{ij} \geq r_C \end{cases}$$

where $r_{ij}$ is the distance between $i$-th and $j$-th particle, $r_c = 2.5\sigma$, $\epsilon$ is the depth of the potential well, $\sigma$ is the effective radius of an atom.

Henceforth the physical properties are expressed in the normalized units. Thus, energy is measured in $\epsilon$, length in $\sigma$, temperature in $\epsilon/k_B$ with $k_B$ being the Boltzmann constant, pressure in $\epsilon/\sigma^3$ and time in $\tau = \sigma\sqrt{m/\epsilon}$ with $m$ being the mass of a particle.

The Newtonian equations of motion were integrated using the velocity form of the Verlet algorithm [21] under periodic boundary conditions and with a time step of 0.005$\tau$.

The simulation technique used in the present work is molecular dynamics in the NVT ensemble, using the velocity scaling thermostat [22], where $V = L^3 \approx 1838\sigma^3$, $L = 12.25\sigma$ is a linear dimension of the system. Thus, the system is characterized by constant density $\rho = N/V \approx 0.75\sigma^{-3}$.

The initial state of the system is prepared an equilibrium liquid at $T_i = 1.3\epsilon/k_B$ above the melting temperature $T_m \approx 0.75\epsilon/k_B$ [23]. Then, the system was cooled to temperature $T_f = 10^{-2}\epsilon/k_B$ at cooling rates $\gamma_1 = 2 \times 10^{-2}\epsilon/k_B\tau$ and $\gamma_2 = 4 \times 10^{-3}\epsilon/k_B\tau$. The number of steps for this procedure was $M_1 = 4 \times 10^4$ and $M_2 = 2 \times 10^6$, respectively. The temperature change was stopped every $M_1/20$ steps. Then, the system was equilibrated in $15 \times 10^3$ steps and radial distribution function (RDF) [24] was calculated in $n = 5 \times 10^3$ steps:

$$g(r) = \frac{2V}{N^2} \left( \sum_{i=1}^{N-1} \sum_{j>i}^{N} \delta(r - r_{ij}) \right),$$

(3)

where $r_{ij} = \mathbf{r}_i - \mathbf{r}_j$ is a distance between $i$-th and $j$-th particles and angle brackets denote $n$ ensembles averaging. If the system is also isotropic the function can be averaged over angles without loss of information, so $g(r) \equiv g(r)$. It should be noted, that $r_{ij}$ in (3) is implicitly dependent on the temperature of the system, therefore $g(r) \equiv g(r, T)$. Then, cooling was resumed after RDF calculation.

3. Regression procedure

The key idea of the problem solution is shown in figure 1. Let fix a value of interparticle distance $r = r_0$. Then temperature dependence of RDF $g(T, r = r_0)$ can be obtained. The next step is to choose an
appropriate regression model that is fitted to data from molecular dynamics simulations. In terms of supervised machine learning, the system temperature is a feature and RDF is a target. This model is used to predict RDF value at the temperature of interest. After repeating this procedure for \( r \) from 0 to \( L/2 \), RDF at desired temperature can be obtained. This RDF needs to be compared with an analogous one calculated with direct molecular dynamics simulations to test the learned model. The obtained RDF, if it is valid, can be used in the further determination of the system configuration.

In this work following regression models are used: linear model [25], \( k \)-nearest neighbors (\( k \)-NN) algorithm [26], decision tree [27]. \( k \)-NN regression is presented in two forms. The first one is based on a fixed number of neighbors. The second one is based on all particles within a fixed radius. Moreover, these algorithms can be uniform with an equal contribution for each neighbor, and weighted by the inverse of their distance.

\[ \text{Enter} \]
\[ r = r_{\text{min}} \]
\[ \text{Consider } g(r, T) \text{ at fixed } r \]
\[ \text{Initialize regression model} \]
\[ \text{Set hyperparameters} \]
\[ \text{Fit model to train data} \]
\[ \text{Predict RDF values at temperatures of interest} \]
\[ \begin{array}{c}
\text{No} \\
\text{Increase } r \\
\text{Yes}
\end{array} \]
\[ r = r_{\text{max}} \]
\[ \text{Construct } g(r, T) \text{ from predicted values for all } r \text{ at fixed } T \]
\[ \text{Exit} \]

Figure 1. Flowchart of the algorithm for radial distribution function prediction.

4. Results and discussions
An example of the described algorithm implementation for the case of \( k \)-NN regression is shown in figure 2. Radial distribution functions at various temperatures and cooling rate \( \gamma_2 \), obtained by classical molecular dynamics simulations as well as regression methods are shown.
Mean-square error (MSE) is used as metrics of models quality:

\[
MSE(r, T) = \frac{1}{n} \sum_{i=1}^{n} [g_i(r, T) - \hat{g}_i(r, T)]^2, \tag{4}
\]

where \( g_i(r, T) \) are observed values, \( \hat{g}_i(r, T) \) are predicted values, \( n \) is the number of samples.

**Figure 2.** Example of the described algorithm implementation for the case of \( k \)-NN regression weighted by the inverse of distances.

A comparison of the mentioned above regression models is shown in figure 3. Their names are displayed on the horizontal axis and maximum MSE values for each of them are displayed on the vertical axis. Regression models that should be noted are \( k \)-NN regression based on a fixed number of neighbors and a decision tree algorithm. These algorithms have minimum MSE almost in all temperature ranges and interparticle distance ones.

**Figure 3.** A comparison of the considered regression models. \( L \) is linear, \( KN \) is \( k \)-NN regression based on a fixed number of neighbors, \( RN \) is variation based on a fixed radius. \( U \) means uniform regression and \( D \) means one weighted by the inverse of their distance. \( DT \) is a decision tree.
5. Acknowledgements

This study is supported by the Kazan Federal University Strategic Academic Leadership Program.

References

[1] Khusnutdinov R M, Mokshin A V and Yulmetev R M 2009 J. Exp. Theor. Phys. 108 417
[2] Khusnutdinoff R M and Mokshin A V 2010 Bull. Russ. Acad. Sci.: Phys. 74 640
[3] Mokshin A V, Chvanova A V and Khusnutdinoff R M 2012 Theor. Math. Phys. 171 541
[4] Mokshin A V and Galimzyanov B N 2014 J. Chem. Phys. 140 024104
[5] Mokshin A V et al. 2015 J. Exp. Theor. Phys. 121 828
[6] Mokshin A V 2015 Theor. Math. Phys. 183 449
[7] Mokshin A V and Galimzyanov B N 2018 J. Phys.: Condens. Matter 30 085102
[8] Khusnutdinoff R M, Galimzyanov B N and Mokshin A V 2018 J. Exp. Theor. Phys. 126 83
[9] Goldman S 1979 J. Phys. Chem. 83 3033
[10] Morsali A, Goharshadi E K, Mansoori G A and Abbaspour M 2005 Chem. Phys. 310 11
[11] Bamdad M, Alavi S, Najafi B and Keshavarzi E 2006 Chem. Phys. 325 554
[12] Matteoli E and Mansoori G A 1995 J. Chem. Phys. 103 4672
[13] Broyles A A, Chung S U and Sahlin H L 1962 J. Chem. Phys. 37 2462
[14] Duh D M and Haymet A D J 1995 J. Chem. Phys. 103 2625
[15] Duh D M and Henderson D 1996 J. Chem. Phys. 104 6742
[16] Choudhury N and Ghosh S K 2002 J. Chem. Phys. 116 8517
[17] Salvalaglio M et al. 2016 J. Chem. Phys. 145 211925
[18] Moradzadeh A and Aluru N R 2019 J. Phys. Chem. Lett. 10 7568
[19] Craven G T, Lubbers N, Barros K and Tretiak S 2020 J. Chem. Phys. 153 104502
[20] Jones J E 1924 Proc. R. Soc. Lond. A 106 441
[21] Swope W C, Andersen H C, Berens P H and Wilson K R 1982 J. Chem. Phys. 76 637
[22] Woodcock L V 1971 Chem. Phys. Lett. 10 257
[23] Smit B 1992 J. Chem. Phys. 96 8639
[24] Rapaport D C 1988 The Art of Molecular Dynamics Simulation (Cambridge: Cambridge University Press)
[25] Yan X and Su X G 2009 Linear regression analysis: theory and computing (Singapore: World Scientific Publishing Co. Pte. Ltd)
[26] Altman N S 1992 Am. Stat. 46 175
[27] Breiman L, Friedman J, Olshen R and Stone C 1984 Classification and Regression Trees (Belmont: Wadsworth)