

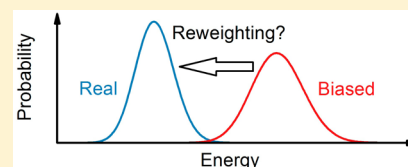
A Comment on the Reweighting Method for Accelerated Molecular Dynamics Simulations

Zhifeng Jing[†] and Huai Sun^{*,†,‡,§}

[†]School of Chemistry and Chemical Engineering and [‡]Ministry of Education Key Laboratory of Scientific and Engineering Computing, Shanghai Jiao Tong University, Shanghai 200240, China

[§]State Key Laboratory of Inorganic Synthesis and Preparative Chemistry, College of Chemistry, Jilin University, Changchun 130012, China

ABSTRACT: Accelerated molecular dynamics (aMD) is a popular method for biomolecular simulations. The estimator to recover the true free energy profile, termed “reweighting”, constitutes a critical part of aMD. Recently the second order cumulant expansion has been recommended as an improved reweighting method. Here we examine the validity of this reweighting method and provide an insight into the need for selecting appropriate boost potentials.



Many enhanced sampling methods rely on modifying the dynamics or the potential energy surface and then recovering the real distribution through reweighting.¹ For methods that require no reaction coordinate, such as accelerated molecular dynamics (aMD),² reweighting involves evaluating the exponential average of potential changes which is inherently noisy. This problem is shared by the calculation of free energy difference from nonequilibrium work.^{3,4} Therefore, many studies have focused on improving the accuracy of reweighting.^{5–11} In a recent Article,⁸ Miao et al. compared several reweighting methods for aMD simulations and concluded that cumulant expansion to the second order is the most accurate, particularly when the boost potential follows a near-Gaussian distribution. We argue that unless the distribution of the boost potential is exactly Gaussian, which is unlikely to be true, the truncated cumulant expansion to the second order may lead to serious deviations. Therefore, the improved reweighting method proposed by the Article in ref 8 should be used with care.

aMD changes the system potential $V(r)$ to a biased potential $V_b(r)$ so that energy barriers can be easily overcome.² This is often achieved by adding a non-negative boost potential $\Delta V(r)$,²

$$V_b(r) = V(r) + \Delta V(r) \quad (1)$$

aMD does not sample the canonical distribution on the unmodified potential, and the reweighting procedure is needed to extract the information about the real system. The biased distribution $p^*(r)$ and the canonical distribution $p(r)$ are related by

$$p(r) \propto p^*(r) e^{\beta \Delta V(r)} \quad (2)$$

where $\beta = 1/k_B T$ is the inverse temperature. Equation 2 provides the foundation for “energetic reweighting”, which is accurate but suffers from large statistical errors.

It has been shown that using the second order cumulant expansion to calculate $\langle e^{\beta \Delta V} \rangle$ can greatly reduce the statistical

noise.⁸ This is equivalent to approximating $p^*(\Delta V)$ by a Gaussian function and obtaining the average through

$$\langle e^{\beta \Delta V} \rangle = \int p^*(\Delta V) e^{\beta \Delta V} d\Delta V \quad (3)$$

The integrand in eq 3 is the un-normalized canonical distribution of ΔV . Although ΔV is not applied in an unbiased simulation, it can be calculated from configurations in any sample distribution. Let ΔV_m and ΔV_m^* denote the most probable boost potentials in unbiased and biased distributions. Expanding $\ln p^*(\Delta V)$ at ΔV_m^* gives

$$\ln p^*(\Delta V) = \ln p^*(\Delta V_m^*) + \frac{1}{2!} \frac{\partial^2 \ln p^*(\Delta V_m^*)}{\partial \Delta V^2} (\Delta V - \Delta V_m^*)^2 + O((\Delta V - \Delta V_m^*)^3) \quad (4)$$

For $\Delta V \approx \Delta V_m^*$, terms beyond the second order in eq 4 can be omitted, and then the second order cumulant expansion is accurate. However, reweighting is mainly controlled by the region where $\Delta V \approx \Delta V_m$, which can differ significantly from ΔV_m^* . Therefore, omitting the high order expansion terms in eq 4 requires $\ln p^*(\Delta V)$ to be quadratic. Miao et al.⁸ used “anharmonicity” to measure the deviation of $p^*(\Delta V)$ from Gaussian and correlated the quantity with the accuracy of cumulant expansion. In fact, likeness of $p^*(\Delta V)$ to Gaussian can be reached as long as the distribution is narrow, but the condition for $\ln p^*(\Delta V)$ to be quadratic is difficult to be satisfied.

The distribution $p^*(\Delta V)$ depends on the entire potential energy surface. To see the relationship, we assume ΔV is a monotonic function of V , as commonly used in aMD.² Consequently, $p^*(\Delta V)$ and $p^*(V)$ are related by

$$p^*(\Delta V) = p^*(V) \left| \frac{\partial \Delta V}{\partial V} \right| \quad (5)$$

Received: March 12, 2015

Published: May 21, 2015



The biased distribution of V is proportional to the degeneracy $g(V)$ and the biased Boltzmann factor,

$$p^*(V) \propto g(V) e^{-\beta(V+\Delta V)} \quad (6)$$

or equivalently

$$\ln p^*(V) = -\beta(V + \Delta V) + k_B^{-1} S(V) + C \quad (7)$$

where $S(V) = k_B \ln g(V)$ is the microcanonical entropy. It can be seen from eqs 5 and 7 that $\Delta V(V)$ and $S(V)$ determine the order of $\ln p^*(\Delta V)$. Generally speaking, $\Delta V(V)$ is a linear⁷ or other analytical function,^{2,12} and $S(V)$ has no simple analytical form.¹³ When $\Delta V(V)$ is linear, only the entropic contribution remains in the second derivative of $\ln p^*(\Delta V)$,

$$\frac{\partial^2 \ln p^*(\Delta V)}{\partial \Delta V^2} = k_B^{-1} \frac{\partial^2 S(V)}{\partial \Delta V^2} = K \frac{\partial^2 S(V)}{\partial V^2} \quad (8)$$

where K is a constant, and eq 4 becomes a second order approximation of $S(V)$. Since $\partial S(V)/\partial V$ does not change sharply even in the presence of a phase transition,¹³ this approximation is appropriate for a small energy range. Nevertheless, it may fail with large boost potentials. When $\Delta V(V)$ is nonlinear, eq 4 would be more complicated, but it does not make the approximation better.

We illustrate the problem with a two-dimensional ($N = 80 \times 80$ lattice sites) Ising model. A linear boost potential $\Delta V = bV$ ($-1 < b < 0$) is used.⁷ Because of the linear relationship, only the entropic term contributes to the second and higher orders of $\ln p^*(\Delta V)$. Figure 1a shows that $\partial S(V)/\partial V$ is nonlinear (i.e.,

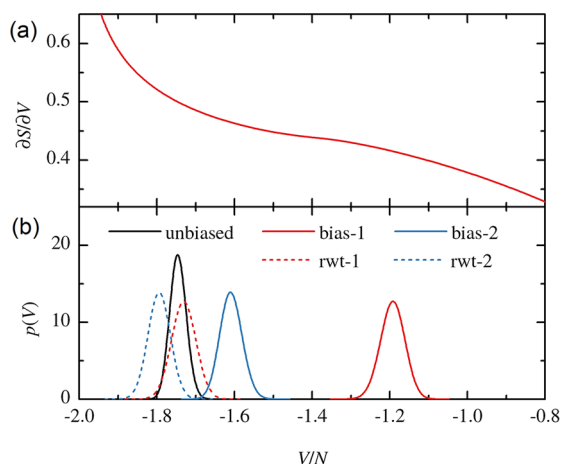


Figure 1. (a) First derivative of microscopic entropy for the 2D Ising model. (b) Second order cumulant expansion reweighting results of potential energy distribution. Solid lines represent unbiased and biased distributions, and dashed lines represent reweighted distributions. Two biased potentials are shown in red and blue.

$S(V)$ is not quadratic) in the energy range. Figure 1b shows the potential energy distributions (PEDs) at $\beta = 0.5$ of unbiased simulation and aMD simulations with two different biased potentials, bias-1 with $b = -0.17$ and bias-2 with $b = -0.07$. The aMD PEDs are reweighted using the second order approximation. The reweighted PEDs clearly differ from the unbiased PEDs despite the fact that the PEDs are near-Gaussian. The bias-1 simulation forces the system to explore a very high energy region; however, reweighting of bias-1 has a more accurate average energy than reweighting of bias-2 due to cancelation of errors. Furthermore, the reweighted PEDs

exhibit the same shape as the biased PEDs. This is because $\partial^2 S(V)/\partial V^2$ is assumed constant in the second order cumulant expansion. Consequently, estimate of energy fluctuation is still biased after reweighting.

The error of the second order cumulant expansion for $\langle e^{\beta \Delta V} \rangle$ can be calculated accurately for the Ising model. The calculated error and the anharmonicity as functions of the average boost potential per lattice site $\langle \Delta V \rangle / N$ are shown in Figure 2. The

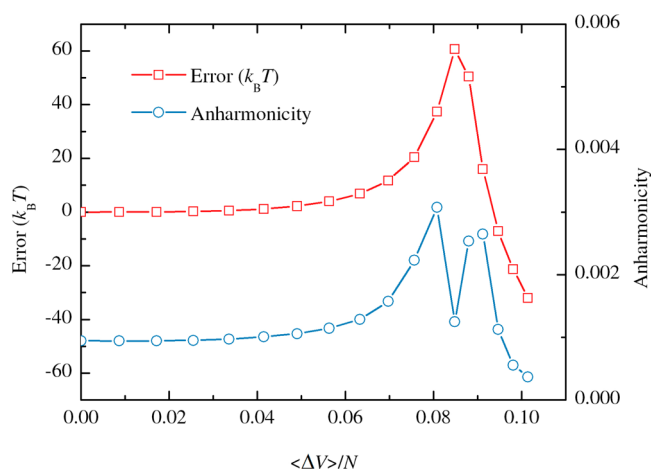


Figure 2. Error of the second order cumulant expansion reweighting and anharmonicity of boost potential distribution at a range of boost potential levels.

error is small for low boost potential but increases sharply when $\langle \Delta V \rangle / N$ goes above 0.05 (comparable to the energy fluctuation). The signed errors range from 60 to $-32 k_B T$ for large boost potentials, while the small error at $\langle \Delta V \rangle / N \approx 0.095$ results from the error cancelation as discussed previously. Despite the large errors, the anharmonicity is very low (less than 0.004) for the entire range of applied boost potentials. More importantly, the trend of the anharmonicity is inconsistent with that of the absolute error. For example, the largest positive and negative errors correspond to relatively low anharmonicity values.

The difficulty arises from the fact that, with the boost potential, aMD may sample an energy region different from the original one, and the microscopic entropy cannot be approximated by a quadratic function. All of these pose a greater challenge in choosing appropriate biased potentials.

AUTHOR INFORMATION

Corresponding Author

*E-mail: huaisun@sjtu.edu.cn.

Funding

We acknowledge financial support from the National Natural Science Foundation of China (Grant Nos. 21073119, 21173146, and 21473112) and the National Basic Research Program of China (Grant No. 2014CB239702).

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

aMD, accelerated molecular dynamics; PED, potential energy distribution

■ REFERENCES

- (1) Christ, C. D.; Mark, A. E.; Van Gunsteren, W. F. Basic ingredients of free energy calculations: A review. *J. Comput. Chem.* **2010**, *31* (8), 1569–1582.
- (2) Hamelberg, D.; Mongan, J.; Mccammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120* (24), 11919–11929.
- (3) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J. Chem. Phys.* **2003**, *119* (6), 3559–3566.
- (4) Gore, J.; Ritort, F.; Bustamante, C. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (22), 12564–12569.
- (5) Shen, T.; Hamelberg, D. A statistical analysis of the precision of reweighting-based simulations. *J. Chem. Phys.* **2008**, *129* (3), 034103.
- (6) Doshi, U.; Hamelberg, D. Improved Statistical Sampling and Accuracy with Accelerated Molecular Dynamics on Rotatable Torsions. *J. Chem. Theory Comput.* **2012**, *8* (11), 4004–4012.
- (7) Sinko, W.; Miao, Y.; De Oliveira, C. A. F.; Mccammon, J. A. Population Based Reweighting of Scaled Molecular Dynamics. *J. Phys. Chem. B* **2013**, *117* (42), 12759–12768.
- (8) Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; Mccammon, J. A. Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J. Chem. Theory Comput.* **2014**, *10* (7), 2677–2689.
- (9) Shirts, M. R.; Pande, V. S. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J. Chem. Phys.* **2005**, *122* (14), No. 144107.
- (10) Kim, S.; Kim, Y. W.; Talkner, P.; Yi, J. Comparison of free-energy estimators and their dependence on dissipated work. *Phys. Rev. E* **2012**, *86* (4), No. 041130.
- (11) Procacci, P. Unbiased free energy estimates in fast non-equilibrium transformations using Gaussian mixtures. *J. Chem. Phys.* **2015**, *142* (15), No. 154117.
- (12) Sinko, W.; De Oliveira, C. a. F.; Pierce, L. C. T.; Mccammon, J. A. Protecting High Energy Barriers: A New Equation to Regulate Boost Energy in Accelerated Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2012**, *8* (1), 17–23.
- (13) Kim, J.; Straub, J. E.; Keyes, T. Statistical-Temperature Monte Carlo and Molecular Dynamics Algorithms. *Phys. Rev. Lett.* **2006**, *97* (5), No. 050601.