Bias in Molecular Dynamics Averages due to Finite Step Size

Stephen D. Bond and Robert D. Skeel

Even if our model for the molecular system is exact, computational resources limit the simulation length, which introduces statistical error when computing ensemble averages. Additionally, molecular dynamics, unlike Monte Carlo, simulations give biased results due to the use of numerical integrators. This applies whether the dynamics is deterministic or stochastic; this discussion considers only the deterministic case.

In molecular dynamics (MD) it is assumed that time averages of "physically meaningful functions" do not depend on a detailed description of the initial conditions, which is a consequence of the ergodic hypothesis. Due to the chaotic nature of the underlying equations of motion, it is impossible to construct an exact trajectory over a long time interval. Hence, it is remarkable that there exist numerical integrators that can approximately preserve the correct invariant density for very long durations, in particular, symplectic integrators such as Verlet.

There are three ways in which temporal discretization introduces bias into averages. First, the finite step size, Δt, modifies the dynamics in a systematic way, with an effect that can be expressed as an asymptotic expansion in powers of Δt. Second, discretization may affect ergodicity [10], which is particularly problematic since it cannot be easily detected or corrected. Third, it introduces energy drift, particularly for nonsymplectic integrators like Nosé-Hoover schemes [2].

The presentation by Stephen Bond addresses the first of these concerns. It describes practical methods for estimating and correcting bias in averages computed from molecular dynamics simulation. Variations of this idea have been explored by several authors including Bond [1, 2, 3], Leimkuhler [2], and Reich [7]. Assuming ergodicity, the error in a MD average can be bounded by the sum of bounds on the statistical and truncation error. The statistical error decays at a rate proportional to $1/\sqrt{\tau}$, where τ is the simulation time. On the other hand, the truncation error is asymptotically proportional to Δt^r, where r is the order of accuracy of the numerical integrator.

The dependence of the computed average on step size is observed only when the simulation is long enough, otherwise statistical error dominates. Bond and Leimkuhler observed up to one percent error in the induced temperature in NVT molecular dynamics simulations based on the Nosé thermostat [2]. Cuendet and van Gunsteren found a much more severe bias in constant pressure simulations due to the use of the virial in the computation of system pressure [4]. In their study, they found that a one percent error in temperature could cause the induced pressure to be off by a factor of 5 to 14.

When a symplectic integrator is applied to a Hamiltonian system, the result is a modified Hamiltonian system which is nearly exactly conserved by the numerical integrator. For example, it can be shown that the Verlet method conserves $H(x,p) = \frac{1}{2}p^{\mathsf{T}}M^{-1}p + U(x)$, to second-order accuracy over long time intervals, where M is a mass matrix and U is the potential energy. However, the modified Hamiltonian $H_{2,\Delta t}(x,p) = H(x,p) + \Delta t^2(\frac{1}{12}p^{\mathsf{T}}M^{-1}U_{xx}M^{-1}p - \frac{1}{24}\nabla U^{\mathsf{T}}M^{-1}\nabla U)$ is preserved to fourth-order accuracy. Estimates of such modified Hamiltonians can be computed at almost no additional cost [1, 9, 5, 8].

The impact of discretization error on MD time averages is to cause the sampling to be from the invariant distribution (or ensemble) of the modified system. Since the effect of discretization error is known, it can be corrected for, resulting in higher-order averages from a single MD simulation. Specifically, samples from the modified distribution can be reweighted to get an estimate from the desired distribution [2, 3]. An alternative approach is to estimate a correction using an analytical estimate of the effect of a Hamiltonian perturbation on time averages [10].

Finally, it might be noted that an estimate of sampling bias is also useful for improving the efficiency of hybrid Monte Carlo methods [6, 8].

References

[1] S.D. Bond. *Numerical Methods for Extended Hamiltonian Systems with Applications in Statistical Mechanics*. PhD thesis, Dept. of Math., Univ. of Kansas, Lawrence, KS, 2000.

[2] S.D. Bond and B.J. Leimkuhler. Molecular dynamics and the accuracy of numerically computed averages. *Acta Numerica*, 16:1–65, 2007.

[3] S.D. Bond. Backward error analysis for microcanonical averages, 2008. In preparation.

[4] M.A. Cuendet and W.F. van Gunsteren. On the calculation of velocity-dependent properties in molecular dynamics simulations using the leapfrog integration algorithm. *J. Chem. Phys.*, 127:184102, 2007.

[5] R.D. Engle, R.D. Skeel, and M. Drees. Monitoring energy drift with shadow Hamiltonians. *J. Comput. Phys.*, 206:432–452, 2005.

[6] J. Izaguirre and S. Hampton. An efficient propagator in phase space of macromolecules. *J. Comput. Phys.*, 200:581–604, 2004.

[7] S. Reich. Backward error analysis for numerical integrators. *SIAM J. Numer. Anal.*, 36:1549–1570, 1999.

[8] E. Akhmatskaya and S. Reich. GSHMC: An efficient method for molecular simulation. *J. Comput. Phys.*, 227:4934–4954, 2008.

[9] R.D. Skeel and D.J. Hardy. Practical construction of modified Hamiltonians. *Siam J. Sci. Comput.*, 23:1172–1188, 2001.

[10] R.D. Skeel. What makes molecular dynamics work? *SIAM J. Sci. Comput.*, in press, 2008.