Based-on-experiences Note on Statistical Molecular Dynamics for Soft Matter Simulations

Rangsiman Ketkaew [*1]

[1]Department of Chemistry, University of Zurich, Switzerland

November 18, 2022

1 Molecular Dynamics

Philosophy: "Doing MD is easy but doing MD correctly is less easy!"

In my opinion, MD is a random process! But we manage it in statistic and scientific ways!

1.1 Equipartition theorem

From wikipedia "*In thermal equilibrium, energy is shared equally among all of its various forms; for example, the average kinetic energy per degree of freedom in translational motion of a molecule should equal that in rotational motion.*"

1.2 Ergodicity

- Ergodic hypothesis: Ensemble average of a property is the same as its time average
  - One long simulation equivalent to many independent short simulations
- An ergodic system spends equal time in equal volumes of the phase space (combinations of positions and momenta)
  - Given enough time, all possible states will be visited
  - There should be no isolated "islands" of phase space disconnected from the rest, no short exactly periodic trajectories

*rangsiman.ketkaew@chem.uzh.ch

2 Initialization

The initial velocities are drawn from a Gaussian distribution with variance

$$\sigma_{i^2} = \frac{k_\mathrm{B}T}{m_i} \tag{1}$$

where k_B denotes Boltzmann's constant, T is the temperature and m_i is the mass of the i^{th} particle.

Thus, the problem boils down to generate random numbers from a Gaussian distribution using uniformly distributed random numbers. This is fortunately quite simple: the Wikipedia article https://en.wikipedia.org/wiki/Box\OT1\textendashMuller_transform shows some very common algorithms how to transform uniform random numbers into Gaussian random numbers.

When we put everything together: every component of the velocity of the i^{th} particle is computed via

$$v_{i,\alpha} = \sqrt{\frac{k_\mathrm{B}T}{m_i}}\,\mathcal{N}(0,1)\,, \quad \alpha \in \{x, y, z\} \tag{2}$$

where $\mathcal{N}(0,1)$ is a Gaussian random number with variance 1 and mean 0.

With this definition, each velocity component follows a Gaussian distribution

$$\pi(v_\alpha)\mathrm{d}v_\alpha \propto \exp\left(-\frac{v_\alpha^2}{2\sigma^2}\right)\mathrm{d}v_\alpha \tag{3}$$

but when you write the distribution of the velocity vector in spherical coordinates and integrate the angular components, you obtain

$$\pi(v)\mathrm{d}v \propto v^2 \exp\left(-\frac{v^2}{2\sigma^2}\right)\mathrm{d}v \tag{4}$$

which is the desired Maxwell-Boltzmann distribution.

3 Ensemble

Ensemble is a representation of the simulation system which is consistent with the real system under experiment condition.

3.1 Microcanonical ensemble

Microcanonical ensemble corresponds to adiabatic process where no transfer of heat between system and environment take place, N, V, and E are fixed. But!! in reality,

system always exchange energy with environment. So this ensemble is meaningless and useless to study the canonical behavior of the system. In experiment we do not control the energy but temperature.

3.2 Canonical ensemble

In canonical ensemble N, V, T are constant. With this ensemble, kinetic energy and temperature are constant, but instantaneous kinetic energy fluctuates.

4 Thermostat

4.1 Velocity Scaling

Valocity scaling is a means of imposing thermostat on the system. It can (easily) be done by multiplying velocity with a scaling factor λ.

4.2 Andersen

It uses random gaussian number (stochastic).

4.3 Berendsen thermostat

This thermostat suppress fluctuation in kinetic energy of high-frequency motion (vibration) to zero-frequency motion such as translation and rotation, resulting in an artifact so-called *flying ice cube*. So this ensemble is very low efficient. Actually this thermostat do not generate any ensembles, neither microcanonical nor canonical ensembles.

As not assigned to any ensemble, Berendsen thermostat is global thermostat.

Advatages

- Stable, able to cope with systems arbitrarily far from equilibrium
- Exponential relaxation (quickly approaches target temperature) with time constant τ

Disadvatages

- Does not generate correct canonical ensemble (fluctuations strongly suppressed)
- Fundamentally incompatible with temperature replica exchange MD
- Does not conserve energy
- Causes severe artifacts when τ is too short ("Flying ice cube effect")

4.4 Bussi-Donadio-Parrinello thermostat

Bussi-Donadio-Parrinello thermostat or canonical sampling through velocity scaling (CSVR) add noise to gives correct fluctuations that are missing/wrong when using Berendsen thermostat.

4.5 Nosé-Hoover thermostat

It is basically an improvement of Nosé thermostat (Original idea by Nosé, reformulated to be more practical by Hoover). It is deterministic method and has non-ergodicity for harmonic oscillator system (proved by the authors) This thermostat explicitly simulates heat bath as an additional degree of freedom ($3N + 1$ DoFs: atomic positions plus heat bath energy).

Advatages

- Correct canonical ensemble, conserves energy of system + heat bath
- Relatively safe (artifact-free) across a range of time constants

Disadvatages

- Only works well for systems relatively close to equilibrium
- Original NH strongly oscillates when disturbed
- Potentially non-ergodic when τ is too short (real issue for original NH, unlikely to matter with chains for typical systems
- Characteristic timescale τ only approximate

4.6 Nosé-Hoover chain

Nosé-Hoover chain (NHC) produces canonical ensemble. It solves a problematic issue of non-ergodicity introduced in Nosé-Hoover thermostat.

4.7 Langevin thermostat

Correct velocity by using random force and constant friction

5 Integrator

"Integrator is the heart of MD."

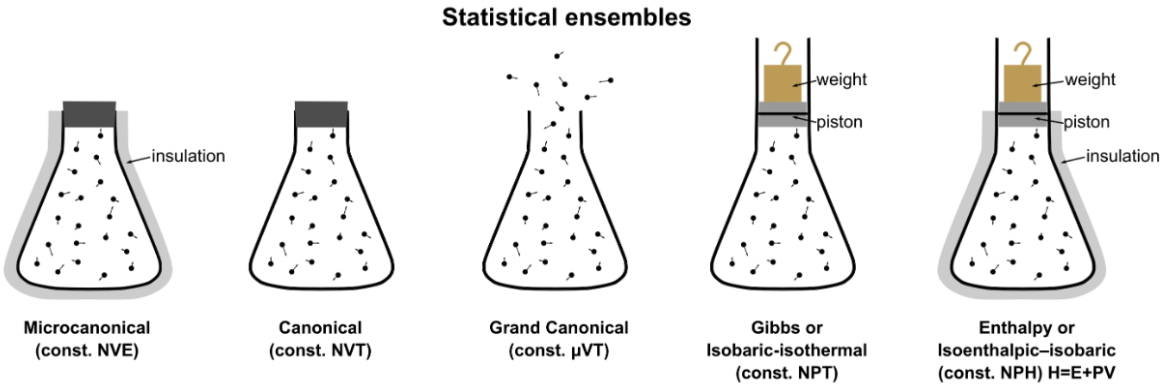
- MD: Numerical integration of the equations of motion
  - Propagating positions x and velocities $\dot{x}$ in time based on accelerations x (forces from chosen potential)
- Unlike general ODE integration methods (Runge-Kutta, ...), MD prefers integrators with special properties (time reversibility = energy conservation).
- One of the efficient integrators is Velocity Verlet integrator
  - Symplectic method: exactly conserves a perturbed "shadow Hamiltonian"
  - Difference between real and shadow Hamiltonian proportional to $O(h^2)$
  - Global error in positions (over a given time interval) also $O(h^2)$

Integration Stability and Accuracy

- Verlet integrator doesn't produce obvious energy drift even with an excessive timestep
  - Some drift due to roundoff errors only observable on $>$ 100 ps timescales
  - Any drift observed in simulations is typically due to inaccurate (noisy) forces
- Symplectic methods become unstable when timestep h $> 0.225T_p$, where T_p is the period of the fastest motion
  - O−H bond: 10 fs
  - C−C bond: 16 fs (triple) - 20 fs (double)
- Reasonable accuracy requires $h \leq 0.1T_p$

6 Problems in MD simulation

6.1 Common statistical error



- Microcanonical ensemble reliable and accurate but not very realistic
- Real-world systems typically in a known temperature and volume/pressure

7 MD simulation

7.1 Setting time constant τ

- Aggressive thermostatting (low) leads to artifacts:
  - Berendsen: flying ice cube effect (energy pumped into low-frequency modes)
  - Nosé-Hoover: potential non-ergodicity, inaccurate integration of heat bath DoFs
- A proper time constant with a reasonable lower bound if strong thermostatting is required
- Any strong thermostatting can disturb natural dynamics (viscosity, diffusivity, ...)
- For equilibrium production simulations, use high enough , comparable with natural relaxation timescales:
  - Liquids: correlation time (2-3 ps for water)
  - Solids: phonon lifetimes (typically > 0.5 ps)

7.2 Thermostatting heterogeneous systems

- Thermostatting a heterogeneous system as a whole can create temperature gradients
- Typical example: "hot solvent - cold solute problem"
  - Big protein in water: force errors on water molecules higher than on protein chain
  - Unequal heating of both parts
  - Shared thermostat drains heat from both components at the same rate, undercooling the protein
  - A good read: https://pubs.acs.org/doi/10.1021/ct8000365
- Solution: use separate thermostats for individual components

7.3 Equilibration protocol

1. Prepare starting geometry (from experiment or static calculation)
2. Optionally run a geometry optimization if starting geometry is very suboptimal
3. Pre-equilibrate near target temperature using NVT with Berendsen and strong coupling (100-500 fs time constant) for 10 ps or more
4. Relax to near equilibrium density in NpT with Berendsen thermostat (τ 1-2 ps) and barostat (τ_p 5 ps for water) for several tens of ps
5. Equilibrate using NHC thermostat and MTK barostat until key properties converge (total and potential energy, density)

6. Alternatively, if preparing for NVT production simulation, equilibrate with Berendsen barostat with a long time constant to reach average density

7.4 Basic equilibrium properties

- Energies, temperature, volume, density
- Radial distribution function
  - Doesn't require long simulations or frequent sampling (few 10s of ps enough)

8 Analysis

Always check energy conservation and kinetic energy fluctuation.

How to check the ergodicity of the system?

- Plot energy v.s. simulation time
- Plot MSD
- Check the velocity autocorrelation function

9 Pitfalls in MD simulations

- Sampling : results obtained from one 100 ns run does not mean much. It is one data point. MD can be used to predict only average behaviour and not a particular process precisely; I think this fact gets lost.
- Forcefields : to all the snooty-atomistic folks, atomistic FFs have issues as well. CHARMM is biased towards helices and OPLS towards random coil, if you are studying protein folding either a)use both or b)use neither.
- Small system to characterize global properties : properties like phase transition/separation are global properties, you can't simulate a small system (e.g., 144 lipids) and claim phase transitions. How do you know it is not a local effect? Sometimes people use PBC to make the system look bigger, and it is not until you get to methods, you realize the size of the system.

Credit: https://www.quora.com/What-are-the-common-pitfalls-in-molecular-dynamics-simulations

10 Summary

- Use thermostats and barostats with care
  - Berendsen methods good for equilibration
  - Prefer Nosé-Hoover thermostats (NHC) and Martyna-Tobias-Klein extended Lagrangian barostat (MTK) for production NpT simulations
  - Do not use short time constants unnecessarily, avoid τ ¡ 100 fs
- Use initial velocities for continuation (restart calculation)
- Monitor basic properties to check for equilibration/convergence

11 References

1. Understanding Molecular Simulation From Algorithms to Applications - D. Frenkel, B. Smit.
2. An Introduction to Statistical Thermodynamics - T. Hill
3. Computer Simulation of Liquids - M. Allen, D. Tildesley