Monte Carlo and MD simulations

What we observe in any system?
• averages of observables (pressure, energy, density)

Given enough time system will visit all states

My observable $A$

\[
\mathcal{A}_{obs} = \frac{1}{b-a} \int_a^b A_t \, dt
\]

\[
\mathcal{A}_{obs} = \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} A_i
\]
Time and space averages

If we believe $\mathcal{A}_{obs} = \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} \mathcal{A}_i$

then

$$\mathcal{A}_{obs} = \sum_{j}^{states} p_j \mathcal{A}_j$$

$\equiv \langle \mathcal{A} \rangle$

and $p_j$ is the probability of state $j$

- $\langle \mathcal{A} \rangle$ is ensemble average and usually $\bar{\mathcal{A}}$ is time average
- if sample with correct probability, we can find $\mathcal{A}_{obs}$
- order of visiting states does not matter
Monte Carlo

How to calculate $\pi$ with random numbers

\[
\frac{\text{points}_{\text{red}}}{\text{points}_{\text{square}}} = \frac{1/4 \pi r^2}{\text{area in square}}
\]

\[
\pi = 4 \frac{\text{points}_{\text{red}}}{\text{points}_{\text{square}}}
\]

while (not converged)

pick random $x, y$

$n_{\text{square}}++$

if ($(x^2 + y^2) < 1$)

$n_{\text{red}}++$

print $4 \frac{n_{\text{red}}}{n_{\text{square}}}$

diagram loosely from Allen, MP and Tildesley, DJ

5/26/2015 [3]
Generating distributions / Monte Carlo

Generating points in a circle? (generating function)

\[ p_{\text{in\_circle}} = \begin{cases} 
1 & x^2 + y^2 \leq 1 \\
0 & x^2 + y^2 > 1 
\end{cases} \]

- we could work out the area of a circle (integrate) by picking random numbers

What does Monte Carlo simulation mean?
- generating points according to some distribution to find an average or integral
- what is our distribution in physical systems?
  - Boltzmann distribution
Monte Carlo and Boltzmann distributions

Boltzmann probability distribution

\[ p_i = \frac{e^{-E_i / kT}}{\sum_j e^{-E_j / kT}} \]

often written as

\[ p_i = \frac{e^{-E_i / kT}}{Z} \]

since we define

\[ Z = \sum_j e^{-E_j / kT} \]

• if we could generate this distribution,
  we could reproduce most properties of a system
• leads to a scheme (not possible)
correct, but not practical scheme

while (not happy)
    generate configuration $r_i$ (conformation of protein, ...)
    calculate $p_i$ (number between 0 and 1)
    generate random number $x$
    if $(x < p_i)$
        accept $r_i$
    else
        reject $r_i$

• result ? a set of $r_i$ with Boltzmann distribution
• problem ? we do not know $\sum_j e^{-\frac{E_j}{kT}}$
We cannot generate points from

\[ p_i = \frac{\frac{-E_i}{e^{kT}}}{\sum_j e^{\frac{-E_j}{kT}}} \]

What if we have two configurations?

\[ \frac{p_i}{p_j} = \frac{e^{\frac{-E_i}{kT}} Z}{e^{\frac{-E_j}{kT}} Z} = e^{\frac{E_j - E_i}{kT}} \]

\[ = e^{\frac{-\Delta E}{kT}} \]
a better scheme

\[
\frac{p_i}{p_j} = e^{\frac{-\Delta E}{kT}}
\]

If we have one configuration to start
• we can work out the relative probability of a second

Convenient convention
• going from old→new  \( \Delta E < 0 \)
  • \( E_{\text{new}} - E_{\text{old}} < 0 \)  energy is better / more negative

Does it matter where you start? What is \( i \) ?
Metropolis Monte Carlo

- generating a distribution \( \frac{p_i}{p_j} = e^{\frac{-\Delta E}{kT}} \)
- if \( \Delta E < 0 \), new is likely (more than 1)
- if \( \Delta E > 0 \), old is \( p_{\text{new}} \) is possible

```
generate starting configuration \( \mathbf{r}_o \)
while (not happy)
    generate \( \mathbf{r}_{\text{new}} \)
calculate \( E_{\text{new}} \) and \( \Delta E \)
if \( \Delta E < 0 \)
    set \( \mathbf{r}_o \) to \( \mathbf{r}_{\text{new}} \)
else
    \( x = \text{rand} \ [0:1] \)
    \( x \leq e^{\frac{-\Delta E}{kT}} \)
    set \( \mathbf{r}_o \) to \( \mathbf{r}_{\text{new}} \)
```

- what if \( \Delta E \) slightly > 0 ?
  - 0.0000000001
- what if \( \Delta E = 10^6 \) ?
- small uphill moves are OK
- bigger moves are less likely
Properties of Monte Carlo

The set of $r_o$ is a valid distribution (ensemble)

- for some property $\mathcal{A}$

\[ \mathcal{A}_{\text{obs}} = \langle \mathcal{A} \rangle = \frac{1}{N_{\text{visited}}} \sum_{i}^{N_{\text{visited}}} \mathcal{A}_i \]

- $\mathcal{A}$ could be density, structural property, $E$, ...
- only works for one temperature $T$

Look at picture.. could I calculate entropy / free energy ?
- for simple systems
MC results (observables / averages)
• only for system at equilibrium
• simulations generate system at equilibrium

What happens for a system out of equilibrium?
• Toy system with 3 states
• for some $T$, at equilibrium
• $p_1 = \frac{5}{8}$, $p_2 = \frac{1}{4}$, $p_3 = \frac{1}{8}$
• if I have 80 copies of the system, most are in state $1$
Reaching equilibrium

System wants \( p_1 = \frac{5}{8} \), \( p_2 = \frac{1}{4} \), \( p_3 = \frac{1}{8} \)

50 : 20 : 10

- start it with 5 : 70 : 5

- all moves 2 → 1 are accepted (large flux)
- the flux from 1 → 2
  - 1 → 2 moves are not always accepted
  - there are less particles in state 1

Moving to equilibrium depends on
- population
- probability
Detailed balance

For any two states (state \(i\) and state \(j\))

Flow \(i \rightarrow j\) must equal \(j \rightarrow i\)

- otherwise?

Flow \(i \rightarrow j\) depends on

- population \(N_i\)
- probability \(\pi(i \rightarrow j)\)

Detailed balance

\[ N_i \pi(i \rightarrow j) = N_j \pi(j \rightarrow i) \]

- detailed balance must apply for any pair \(i, j\)

all textbooks use \(\pi\) for probability here
Ergodic

Assumptions

• I can do integrals because
  • I will visit every state
  • I can calculate $p_i$ for all states
• I will visit every state

Alternatively

For any $i, j$

• $\pi(i \to j) > 0$
• may require a finite number of steps: $i \to k \to m \to j$
• must be satisfied
version 1
• decide on $r_{\text{max}}$
• pick a particle at random
• pick random $\Delta x, \Delta y, \Delta z$
  
  \[ 0 < \Delta a < r_{\text{max}} \]
• apply move
• accept / reject move

version 2
• decide on smaller $r_{\text{max}}$
• foreach particle
  • pick random $\Delta x, \Delta y, \Delta z$
    \[ 0 < \Delta a < r_{\text{max}} \]
• apply move
• accept / reject
Moves

• both kinds of move OK
• note
  • "accept / reject"
More generally,
• how big is $r_{max}$?
• big
  • system moves faster
  • more moves rejected
What if my particles are not spheres?
• rotations also necessary

• time has no meaning
Bonded systems

Protein (lipid, polymer, ..)
Random $\Delta x$?
- nearly all will stretch a bond
  - high energy: rejected move
- only feasible method
  - random rotations $\Delta \theta$

In general
- most kinds of simple moves OK
- must maintain detailed balance, ergodicity
- question of efficiency
  - high rejection rate means lots of wasted calculations
More moves – *N* particles

\[ \frac{p_{\text{new}}}{p_{\text{old}}} = e^{-\Delta E/kT} \]

I have defined temperature

- and \( N \) particles and \( V \)
- called NVT simulation

Could I have varied something else?

- what if I tried to put particles in / take out?
  - sometimes energy ↑sometimes↓
- system will fluctuate around \( \langle N \rangle \)
- this would not be NVT
Periodic Boundary Conditions

Technical point relevant to gases, proteins in water...

Behaves like an infinite system
Infinite interactions?

Neighbours of blue particle

- only use the nearer
- not really an infinite system
- volume defined by box
Problems with Monte Carlo

while (not happy)
  propose move
  accept / reject move

Small steps?
• system moves slowly: long time to visit all states

Big steps?
• calculate energy
• reject move
  • no progress, wastes time
Dense Systems and Monte Carlo

Random moves?
• most moves rejected

Dense systems?
• liquids
• proteins, polymers, ...

Solutions
• cleverer MC moves (later)
• MD
Why do molecular dynamics simulations?

Real world
- box of gas, molecule in space, protein molecule in water
- atoms hit each other,
  - share energy, box expands/contracts,..
  - soon reaches equilibrium
- visits low energies (often), high energies (less often)
- visits entropically favoured regions
- we stick in a thermometer
- measure density, ...

What have the atoms done?
- feel forces and move
- an MD simulation just copies this
What do we expect? Molecular Dynamics

one particle in a well

\[ \vec{F} = -\frac{\partial U}{\partial \vec{r}} \]

Unlike MC, particles have kinetic energy \( E_{\text{kin}} \)
Kinetic and potential energy

Our system is isolated (no work done)

$E_{tot}$ never changes

- conserves energy (no work done on system)

$$E_{tot} = E_{pot} + E_{kin}$$

For one particle $E_{tot} = E_{pot} + E_{kin} =$ constant
Lots of particles

Particles hitting each other
  • exchanging energy

Total system
  • conserves energy

One particle?
  • maybe at bottom but moving slow ($E_{\text{kin}} + E_{\text{pot}}$ small)
  • per particle energy no longer conserved (may gain or lose)

Many particles
  • distribution of velocities
  • distribution of potential energies
Boltzmann distribution in real world

One version of real world (N, V, T)

• constant number of particles, volume, temperature
• today \( E = E_{kin} + E_{pot} \)
• \( Z \) is partition function
• earlier \( Z = \sum_i e^{-\Delta E_i / kT} \)

But now we have kinetic energy \( E_{kin}(p) \)

• where \( p = m\dot{x} \)
  • potential energy \( E_{pot}(r) \)
• if we write in continuous form ...
Partition function for MD

Usually write $\mathcal{H}(\mathbf{p}, \mathbf{r}) = E_{\text{kin}}(\mathbf{p}) + E_{\text{pot}}(\mathbf{r})$

- "Hamiltonian"

All the states are defined by all possible momenta and coordinates

- sum over these: $Z(N, V, T) \propto \int d\mathbf{p} \int d\mathbf{r} \ e^{-\frac{\mathcal{H}(\mathbf{p}, \mathbf{r})}{kT}}$

often see $H(\mathbf{p}, \mathbf{r})$ or $\mathcal{H}(\Gamma)$
MD Method

For any particle we can calculate forces

Newton's law

\[ F = ma \] often better written \[ \ddot{x} = \frac{F}{m} \]

If we know acceleration

• we can get velocity
  from velocity
• can get coordinates

\[ \hat{F} = -\frac{\partial U}{\partial \hat{r}} \]

while \( nstep < \text{max\_step} \)

calculate forces
integrate to get new coordinates

\{ averaging, sampling, ...

nstep ++
Starting system

Initial coordinates
- protein model
- protein from protein data bank (PDB)
- protein + proposed ligand
- box of liquid

Do initial coordinates matter?
- in principle: no
  - infinitely long simulation visits all configurations, reaches equilibrium
- in practice: yes
  - bad examples
    - no simulation is long enough to predict protein conformation
  - take water configuration and run at ice temperature
Initial velocities

First consider temperature – reflects kinetic energy

\[
\left\langle \frac{1}{2} m v^2 \right\rangle = \frac{1}{2} kT
\]

where \( v^2 \) could be \( v_x, v_y, v_z \)

leads to definition

\[
T(t) = \sum_{i=1}^{N} \frac{m_i v^2(t)}{kN_f}
\]

- where \( N_f \) is number degrees of freedom \( \approx 3N \)
- we could use this to get initial velocities \( \langle v^2 \rangle = \frac{kT}{m} \)
Initial velocities

Would one $\langle v^2 \rangle$ be OK?

- not very good
  - $E_{\text{kin}}$ correlated with $E_{\text{pot}}$

Either

- use more sophisticated distribution
- do not worry
  - system will go to equilibrium
    - velocities will reach sensible values
Getting new velocities / coordinates

constant acceleration

\[ x_t = x_0 + vt + \frac{1}{2}at^2 \]

or

\[ x_t = x_0 + \dot{x}t + \frac{1}{2}\ddot{x}t^2 \]

OK for constant acceleration

- try to use formula to predict future time

big \( \Delta t \) / step
big error

small \( \Delta t \) / step
small error
slow
Fundamental problem with integration

- We want to use big $\Delta t$ (speed)
- We must use small $\Delta t$ (accuracy)

All $\Delta t$ will give us some error
- numerical integration is never perfect

How small is $\Delta t$?
- depends on fastest frequency / steepest walls in energy
  - usually bonds
- for proteins at room temperature
  - $\Delta t \approx 1$ fs (femtosecond $10^{-15}$ s)
- high temperature $\Delta t$ should be smaller
Noise and heating

General rule

- noise heats the system
- formally difficult to prove
- $E_{kin} = \frac{1}{2} mv^2$

- no kinetic energy
- $E_{kin}$ due to noise

extra velocity
Noise-free Simulation

Energy conservation: Absolute rule $E_{pot} = f(r)$

- no time component
- invariant under translation, rotation

When violated?
- $(r)$ does not change, but $E_{pot}$ changes: $E_{tot}$ changes
Noise Sources

Integrator

- coordinates do not match velocity
  \[ E_{\text{kin}} \text{ wrong: } (E_{\text{kin}} + E_{\text{pot}}) \neq \text{constant} \]
- energy not conserved

Numerical noise

- \( E_{\text{pot}} = f(r) \)
- initial coordinates \( (r) \) quoted to 3 decimal places
- really less accurate

Cutoffs

- within cutoff rotation restricted
- outside cutoff rotation suddenly free

Result

- heating
Equilibrium

Remember MC story
- system not at equilibrium? eventually equilibrates

MD
- start in high energy $E_{pot}$
- $E_{pot}$ converted to $E_{kin}$

Some high energy conformation
- relaxes
- $E_{pot}$ converted to $E_{kin}$

MD system will not
- really find low energy
- known temperature
MD in a closed system

• An isolated molecule should not lose energy
• A repeated box will not lose energy
• Formally system is
  • NVE (constant $N_{\text{particles}}, \text{volume, energy}$)

Problems
• we want to set the temperature of the system
• we may have noise / heat creating energy

Cure
• thermostat
Bath

imagine infinite bath at desired temperature
  • heat will flow in or out
  • at equilibrium no flow of heat
    • maybe removal of noise/heat
How to implement? Many ways
  Occasionally:
  1. introduce a fake particle desired temperature / collide
  2. pick a particle at random / give average $v$ for temperature
  3. Easy method – weak coupling...
Weak Coupling

Remember temperature* \( E_{\text{kin}} = \sum_i^N \frac{1}{2} m_i v_i^2 = \frac{3}{2} NkT \)

Goal: heat leaves system depending on how wrong temperature is
\[
\frac{dT(t)}{dt} = \frac{T_0 - T(t)}{\tau_T}
\]

- \( T_0 \) is reference temperature
- \( \tau_t \) is a coupling / relaxation constant
  - \( \tau_t \) tiny, heat moves fast. \( \tau_t \) big, ...
- to implement this idea? Multiply velocities

*Slight simplification of formula

Implementation of weak coupling

Scale velocities, \( v_{new} = \lambda v_{old} \) and \( \lambda = \left(1 + \frac{\Delta t}{\tau_T} \left( \frac{T_0}{T} - 1 \right) \right)^{1/2} \)

Intuitively

- \( \Delta t \) (time step) big ? temperature will change more
- what if \( T_0 = T \) ?
- square root ?
  - wrong \( T \) reflects a difference in \( v^2 \)
Importance of heat baths

Does not conserve energy

In principle
• bring a system to equilibrium for temperature

In practice
• avoid damage due to numerical errors / approximations

For a system at equilibrium
• heat bath should do nothing

Does allow artificial tricks
• gently heat a system and watch behaviour
• gently cool a system and "anneal" it (more later)

Extension to other properties
• analogous reasoning for pressure bath

Andrew Torda
26/05/2015 [44]
# dynamics versus Monte Carlo

<table>
<thead>
<tr>
<th>MC</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>any cost/energy OK</td>
<td>requires continuous $E_{pot}(r)$</td>
</tr>
<tr>
<td>time usually invalid</td>
<td>gives time scales</td>
</tr>
<tr>
<td>most moves OK</td>
<td>physical trajectories</td>
</tr>
<tr>
<td>temperature from acceptance/rejection</td>
<td>has explicit $E_{kin}$ and temperature bath</td>
</tr>
<tr>
<td>easy to program</td>
<td>can be difficult</td>
</tr>
<tr>
<td>both yield a Boltzmann distribution</td>
<td></td>
</tr>
<tr>
<td>both include entropy</td>
<td></td>
</tr>
</tbody>
</table>