

# Stochastic Methods wrap-up

10.34 Dec. 7, 2015

# Multidimensional Integrals

- Many important multidimensional integrals have integrands of this form:  $p(\underline{x})f(\underline{x})$  where  $p$  is a probability density.

- Often an un-normalized  $p$  is known:

$$p(\underline{x}) = w(\underline{x}) / \iiint w(\underline{x}) d\underline{x}$$

- Boltzmann distribution:  $w(\underline{x}) = \exp(-E(\underline{x})/kT)$

- Bayesian analysis of experiments:

$$w(\underline{\theta} | \text{data}) = p_{\text{prior}}(\underline{\theta}) \exp(-\chi^2(\underline{\theta}, \text{data})/2)$$

These are candidates for Metropolis Monte Carlo

# Metropolis Monte Carlo challenges

- Selecting step length in each dimension
  - want it same order of magnitude as the width of the high-w region
  - If too large, very few steps accepted
  - If too small, takes large number of steps to cover the high-w region.
- Achieving high accuracy requires very large  $N$ 
  - adding one more sig fig requires 100x more samples than used so far.

# Master Equations

- Often we have time-dependent probability of the system being in a discrete state:  $P_{\underline{N}}(t)$ .

$$d\underline{P}/dt = M * \underline{P} \quad \underline{P}(t_0) = \underline{P}_0$$

- Number of possible states usually enormous
  - e.g. problem 2 catalysis: 100 sites, each could be in any of 4 states:  $4^{100} \sim 10^{60}$  distinct states
  - M is even bigger  $(4^{100})^2$
- Different values of all these P's at each time step!

*So usually we cannot compute all the elements of  $P$ , and we cannot sample even a small fraction of all the possible states  $N$*

- Common approach: sample from  $\underline{P}(t)$  using Kinetic Monte Carlo (Gillespie algorithm).
  - Each trajectory requires computing  $2*(t_f-t_o)/\Delta t$  random numbers.
  - Low-probability states will not be sampled at all.
  - $N^{-1/2}$  scaling: hard to achieve high precision
- Decide on which quantities  $\langle f \rangle$  you are trying to compute before you start
  - Can compute several at same time “on the fly”.
  - May not be able to store all the trajectories.

# What is the initial condition $P_0$ ?

- Often have an idea of  $\langle N \rangle$  but not full probability distribution of  $N$ .
- Common approach: assume  $N_A$ ,  $N_B$ , etc. uncorrelated initially, and use a separate Poisson distribution for each one:

$$P_{\text{poisson}}(N) = \langle N \rangle^N \exp(-\langle N \rangle) / N!$$

$$P_0(N_A, N_B, \dots) = P_{\text{poisson}}(N_A) * P_{\text{poisson}}(N_B) * \dots$$

When you start each KMC trajectory, you need to sample from a Poisson distribution for  $N_A$ ,  $N_B$ , etc.

# Which processes to include in Kinetic Monte Carlo?

- Accelerate calculation by making  $\Delta t$  as long as possible, i.e. omit very fast processes (which will be “equilibrated” on long time scale).
- The low-probability processes are not going to be sampled adequately, so consider omitting them completely if it speeds calculation.
- “Adequate sampling”: statistical sampling errors  $\sim \sqrt{\text{number of samples with positive result}}$ .

# For example, in Problem 2....

- If the main reaction  $A+B$  is very fast, then you will get a lot of trajectories showing this reaction, and can get pretty good sampling statistics.
- On the other hand, if the coking reaction is slow, you may not sample enough trajectories that show a coking event to be able to reach any conclusions about it.
- If diffusion is too fast,  $\Delta t$  will be very small and it will be very expensive to compute a trajectory.
  - Sometimes people assume diffusion is equilibrated and so create a different random distribution of A and B on surface at each time step corresponding to (slow) reactions

# Molecular Dynamics

- What is it: Solving motions of atoms or clumps of atoms using Newton's equations of motion.
  - Usually fitted force fields but can be from DFT
  - Can include quantum effects e.g. by RPMD
  - Typically use Velocity Verlet algorithm to integrate the ODEs (conserves energy to high accuracy)
  - Typically use a “thermostat” to capture energy fluctuations due to contact with a thermal bath
- Alternative to Metropolis Monte Carlo for steady-state properties of molecules
- Can be used for computing time-dependent properties (it is a more-or-less exact simulation of what the molecules are really doing).

# Limitations of Molecular Dynamics

- For molecules the stretching vibrations are very fast, so need to use very small  $\Delta t$
- This usually limits  $(t_f - t_o)$  to nanoseconds
- So only can determine static quantities which have equilibrated faster than nanoseconds
- Only can follow dynamic processes which occur in nanoseconds.
- Issue with sampling over large number of possible initial conditions.

MIT OpenCourseWare  
<https://ocw.mit.edu>

10.34 Numerical Methods Applied to Chemical Engineering  
Fall 2015

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.