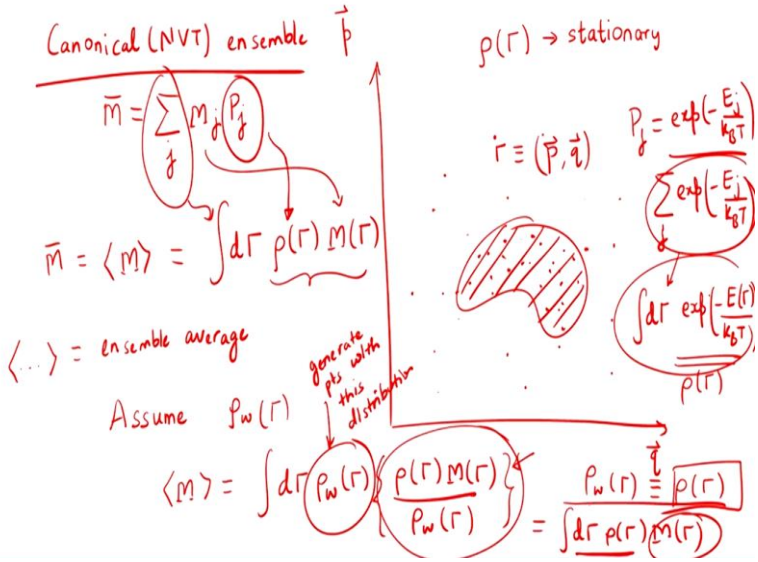
Advanced Thermodynamics and Molecular Simulations
Prof. Prateek Kumar Jha
Department of Chemical Engineering
Indian Institute of Technology, Roorkee

Lecture - 34
Markov Chain Algorithm, Condition for Equilibrium and Detailed Balance

Hello all of you, in the last lecture we have been discussing the basics of the Monte Carlo scheme, how it can be a method to sample the phase space and how Monte Carlo method that is actually a method for numerical integration can be important for thermodynamics. So, in today's lecture we will take the discussion a bit further and try to develop a workable algorithm for a Monte Carlo simulation.

(Refer Slide Time: 01:02)



So, start with, I will do the same phase space that I have been doing in the last few classes we are interested in sampling this particular phase space p and q that is momenta and generalized coordinates and Monte Carlo really works if we are interested in the equilibrium behavior, because it is then the density or the probability density of the states. So, all of these are different states or phase points, so the probability density becomes stationary, that is it becomes independent of time and that is the equilibrium state where the Monte Carlo ideally works and it is in that state that instead of following any particular state point we can simply sample different points in the phase space and get the properties of interest.

So, going further if I want to determine any evidence property in any of the ensembles. So, I have been discussing the canonical ensemble, so let us say if I am interested in computing any evidence property in that ensemble. We have discussed that it is given by something like-

$$\bar{M} = \sum_j M_j P_j$$

where p j is the probability of that state.

So, in here we are typically talking about, so as to speak a macro state in there, because many possible p and q values can give rise to the same energy values. But in general description since I was telling you that in classical mechanics we are talking about a continuous phase space since the number of possible phase points are very very large equally endless we can assume that it is a continuous space and therefore I can replace this summation by an integration similarly this can be replaced by an integration over the p and q or in general an integration over the Γ that is the representation of the p and q together.

So, in general then I can write m bars we prefer the representation <m> under brackets where this thing corresponds to an ensemble average. So, that the meaning of the average is kind of more clear, we are already always referring to some sort of an ensemble on which the average is being conducted, when we are doing it numerically using Monte Carlo or molecular dynamics, of course we are not doing the entire theoretically possible ensemble but a subset of that. So, the ensemble level is computed over whatever I would say reduced ensemble or whatever part of the ensemble that we had sampled. So, this M, this average value of m can then be represented in continuous space as something like-

$$\bar{M} = < M > = \int d\Gamma \, \rho\Gamma \, m\Gamma$$

and why is that, because I am replacing m_j by m of Γ and replacing the property values at different energy states by property value as a function of Γ because in continuous limit it becomes a continuous function the probabilities in discrete space are replaced by the probability density in continuous space and the summation is then replaced by the integration.

So, it is this integration that I want a Monte Carlo simulation to perform and towards the end of the last lecture I was telling you that, of course we can do random sampling in the phase space, but that is probably not the way to go because this entire integrand here can take very large value in some part of the phase space and take a small value in other parts of phase space.

So, whenever we are sampling it since our computational efficiency of sampling is limited to how many samples we are doing, it is better to sample more in the regions of interest, just for the sake of illustration let us say for example, this is my reason of interest that is where this integrand has a high value. So, whenever I am sampling it I will tend to sample more in the region inside this which is having a high value of the integrand and I will tend to sample less in other regions so because we want to get the correct value of I would say as accurate value of the integrand integral as possible and therefore since the larger values of integrands contribute more to the integral, this will at least get the large values right and smaller values anyway are small, so they will contribute less in there and since it is an exponential function in there if you may recall there was a Boltzmann factor which has exponential functions so they decay pretty fast as an exponential and therefore what is found generally is for a majority of phase space the integrand is almost negligible only in some part of the phase space it is quite significant.

So, that is the reason why we need to have some sort of an important sampling, so the basic idea in the important sampling is the following. So, we assume some weighting function and let us say I want to call the waiting function by some ρ_w and this need not be the same as the probability density that we have, but it can be like any function of the phase space. So, then I will compute M as

$$< M > = \int d\Gamma \, \rho_w \Gamma \, \frac{\rho(\Gamma) M(\Gamma)}{\rho_w(\Gamma)}$$

So, effectively what we would be doing in that case is we will generate points with this distribution, it is not going to be a uniform distribution we will generate points using this distribution and once we generate that we compute this quantity under the curly brackets and by taking an average of that quantity given that the points have been generated using ρ_w Ω, $\rho_w \Gamma$ we can get the average value for the property at the interest.

So, now a clear choice you might imagine that I can take $\rho_w \Gamma = \rho \Gamma$ that is same as the probability density that is present and if we are doing that then we are pretty much sampling according to the ρΓ distribution and in that case i will be simply be averaging over the MΓvalue. So, when I was doing with ρw in that case I am sampling with these properties computed at every point and I am averaging over that, when I take ρ_w to be equal to ρ in that case I will average simply the M values over that. So, clearly when I keep the waiting density

ρ_w to be same as ρ, it is clearly more convenient, now no matter how we are doing it the key question is that do we really know this quantity ρ(Γ) and the answer is no because this ρ in fact is coming from the probability of states and the probability of states themselves where something like-

$$P_j = \frac{\exp\left(-\frac{E_j}{k_B T}\right)}{\sum_j \exp(-\frac{E_j}{k_B T})}$$

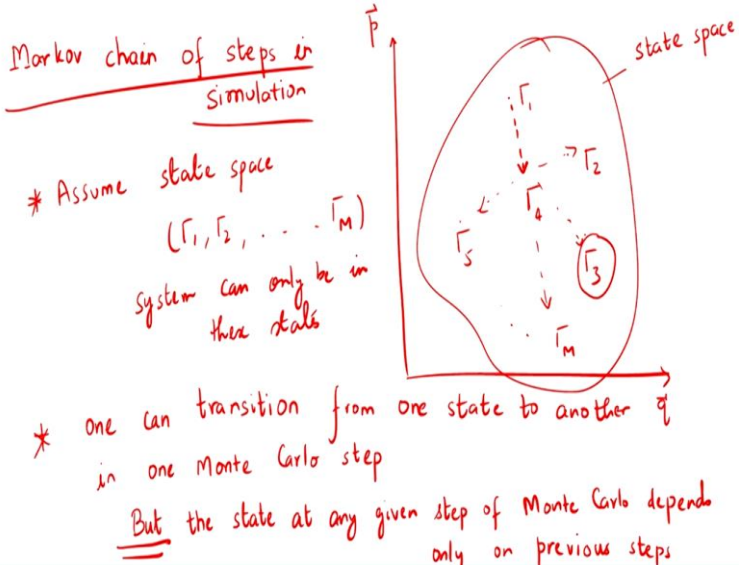
where the denominator is actually the partition function in continuous space the partition function can be written as something like-

$$\sum_j \exp(-\frac{E_j}{k_B T}) = \int d\Gamma \exp\left(-\frac{E(\Gamma)}{k_B T}\right) = \rho\Gamma$$

So this is essentially something like my ρΓ in that particular case and I do not really know this particular integral because if I already know that why exactly are we using Monte Carlo simulation, because ultimately if I know the partition function, I have already said that I know everything about the system. I can compute its derivative find the properties of interest as a derivative of that because partition function is generating function and then I am pretty much done, so the whole idea was to obtain the thermodynamic behavior and so far the argument is like you have to give me the partition function, that is clearly not the way forward.

So, is there any alternative and the answer is yes there is an alternative and that is a very interesting I would say a mathematical trick that works here. So, instead of computing the partition function in Monte Carlo simulation, we directly compute the average properties by this particular trick and the way it works like is the following.

(Refer Slide Time: 12:00)



So, what we do is we assume what is known as a Markov chain of steps in our simulation and this means actually two things first of all I assume that there are certain number of states that I am going to sample and let me call those states Γ_1, Γ_2, Γ_3, Γ_4, Γ_5 to some Γ_m and of course m is going to be very large there is no particular order that I want to represent here. I just want to mention that there are certain number of states that I am going to sample and these constitute what is known as the state space and then that is the first assumption, so we assume a state space and system can only be in these states.

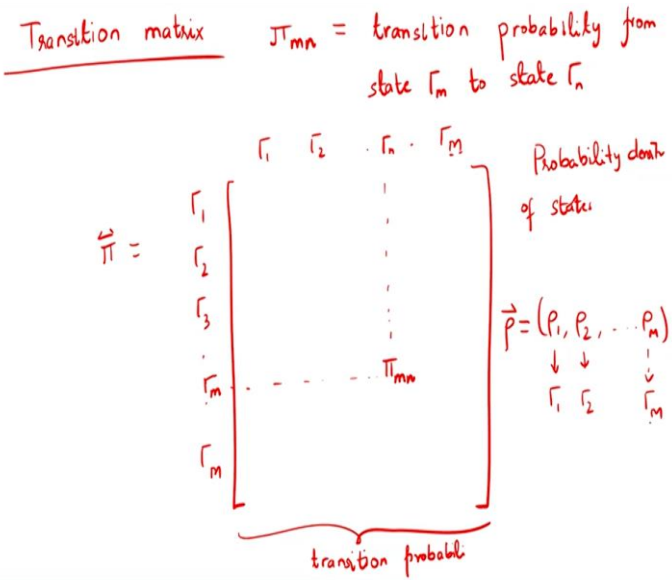
So, in some sense we can therefore say that the state space is like the ensemble, I will do keep in mind that since we are doing some kind of a numerical simulation we are not counting all possible states. So, we will be counting only a small subset of that, so that has to be kept in mind that whenever we are doing a Monte Carlo simulation it is not that we are doing over an infinitely many possible states or all possible states of a system some possible states that we want to explore and that has to be a good representation of the ensemble that comes later. So, the number of states or the value of M should be very large and so on but let us not get there let us assume that there are certain number of states there.

Now the next thing we assume is we one can transition from one state to another in one Monte Carlo step and whenever we are saying that in general, we are saying that we can go from any state in the state space to any other state in the state space but the state at any given step of Monte Carlo depends only on the previous step just to give you one example. So, let us say in one Monte Carlo step, I transition from Γ_1 to Γ_4 and this may happen with certain probabilities we will get to that. Now from Γ_4, I can go pretty much anywhere in the next step. So, let us say

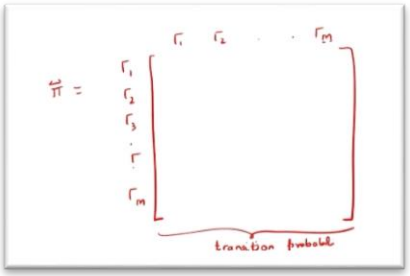
in second step I get to Γ_3. So, if I am at Γ_3 now I can say with certainty that I came from Γ_4, I transitioned from there. But I cannot say what happened prior to that. So, the memory of the system is only of the previous step, so at any given time the state is determined only by where I was last.

One example of I would say one simple example of a Markov chain is a for example a random walk. So, let us say for example if I am walking left and right, so where I am right now it clearly depends on where I was in the last step, because from there I took a left or a right but it does not quite depend on where I was 2 steps back, 4 steps back, 8 steps back and so on. So, decision is made at every step of the Monte Carlo simulation there is no memory so as to speak when we go to more than one steps, that really is somewhat of a difficult concept to think of but the idea here is that for thermodynamic purposes the probability of the transitions will only depend on the previous step and the current step. It does not depend on anything else that where I was, what has happened there and where I want to go after several number of steps and so on. Only the last step determines the current state only the current state will determine the future states, the previous states will not affect the future states only the current state will affect the future states only the last step affects the current state of the system and so on.

(Refer Slide Time: 18:20)



So, with these two ideas I can then define something known as a transition matrix π_{mn}, where π_{mn} is the transition probability from state Γ_m to state Γ_n and just to think about it we can think about it as some sort of a matrix, because in general there can be a large number of states to begin with that is my state space and starting from any of that I can go to any other state and into this brackets here I will populate the transition probabilities.

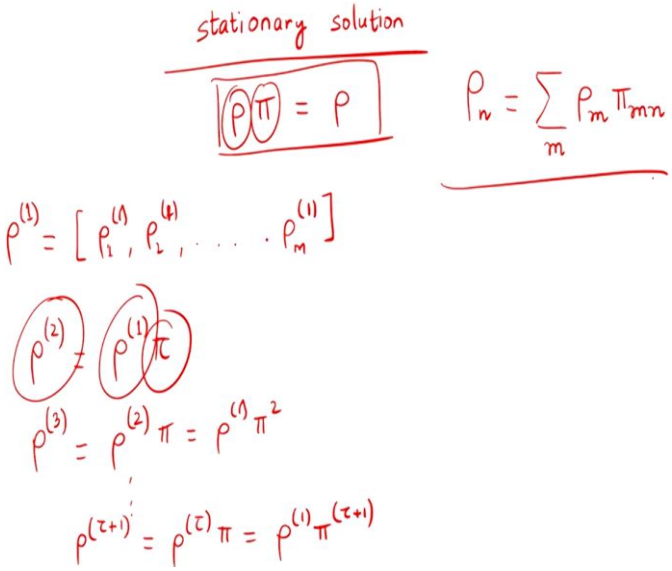


So, let us say for example for Γ_m here and Γ_n here the corresponding element will represent the probability to go from state m to state n that is π_{mn}. So, now I can also write the probability density of states as something like-

$$probability\ density\ of\ state\ (\vec{\rho}) = (\rho_1, \rho_2, \ldots\ldots\ldots\ldots\ldots\ldots\ldots \rho_m)$$

where ρ_1 is the probability density to be at Γ_1, ρ_2 is the probability density to be at Γ_2 and ρ_m is for the Γ_m. So, now what we can then say and until so far we have not made any assumption regarding ρ except that it has to be there, something has to be assumed.

(Refer Slide Time: 21:00)



So, now we can argue that once we reach a stationary solution, what is going to happen is my rho multiplied with π will become equal to ρ and the reason for that is the following. So, let us say for example I start with some probability density ρ1, so you can talk about probabilities as well since it is discrete state but I have used ρ for the density. So, let us keep on that notation but keep in mind that this can also be discrete probabilities.

So, let us say if I start with some probabilities ρ1 of the states. So, if I start from there I know that I have some value of the probability density for all the states that I am considering and this is the some sort of an initial distribution of the system among various states and many of the probabilities can be like even 0, if there is no system in that particular state. So, what will happen after one step?

So, after one step what will happen is the new distribution will be ρ1 multiplied with my transition probability π, because π is telling me the probability of going from state m to state n. So, if I multiply the probability density vector with the transition probability matrix what I will get is the new probability density vector. So, if I want to then compute in the next step it is going to be the ρ in the previous step π and that is equal to ρ1 π^2 and I keep on doing it then after some τ plus one step it is going to be-

$$\rho^{\tau+1} = \rho^{\tau}\pi = \rho^{1}\pi^{\tau+1}$$

I can keep on doing that and the probability density at the states will keep on evolving as the Monte Carlo simulation proceeds. Now this means that my probability density is changing as the Monte Carlo is going on, keep in mind here is that there is no time here so as to speak. So, I can simply defined transition from one state to the other, but I did not mean to say that one step correspond to some certain time, it is simply a hypothetical sort of a step, there is no real dynamics that is taking place in this kind of situation.

So, then what exactly a stationary solution means. A stationary solution means when the probability density is not changing and what is that, that will happen when the ρ multiplied by π that is the probability density in the following step becomes same as ρ that is the density in the current step. So, ρ should eventually converge to a stationary solution because as long as that does not happen we cannot say that we are at equilibrium, because equilibrium is defined as the stationary state that is when the probability densities do not change and therefore the condition to get that is simply ρπ = ρ and this has been done without any consideration to the actual mathematical form of the probability densities without much consideration to even the partition function and so on this is simply I would say a result that has come by construction whenever we are doing a Markov chain.

So, the key assumption here is that at any given step let us say step two the probability density only depends on the previous step that is ρ1 multiplied by the transition probability and this π is a kind of a constant matrix that does not depend on the probability density so as to speak.

So, with this sort of an idea we can write this so ρ keep in mind is a vector and π is a matrix, so we can write this as something like the nth element of ρ is equal to-

$$\rho_n = \sum_m \rho_m \, \pi_{mn}$$

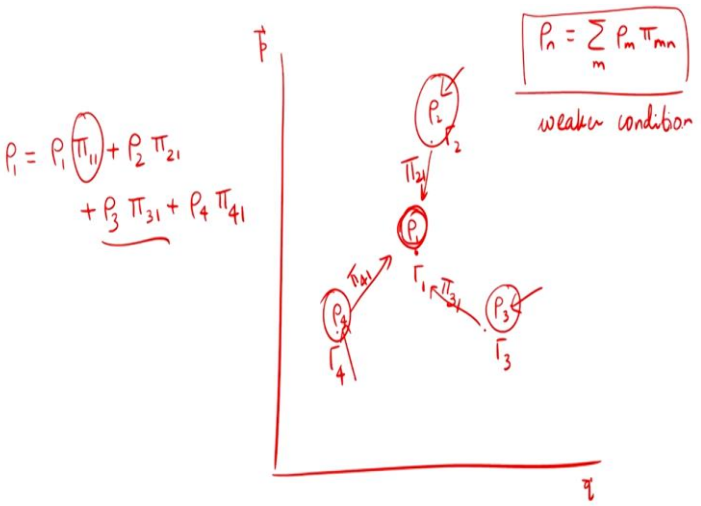
So, that really comes from the multiplication of the ρ vector with the π matrix, so ρ vector is actually a ρ vector that has m columns and the π matrix has m rows. So, the new matrix will again be a ρ vector and every term will be given by this, you can do the matrix algebra here and find that it is true. There is another way to show that what exactly it means, so all it means is that when we reach the stationary state in this kind of a picture what should happen- is that since at every state we are coming from some other state, so if I want to find the probability density in that state is it has to be sum over the probability density of other states multiplied by the transition probability. So, let me put that in a slightly better way in the plot.

So, let us say if I am looking at this Γ_1 here and let us say there are these three other state Γ_2, Γ_3 and Γ_4 and the probability density for these ones are ρ_1, ρ_2, ρ_3 and ρ_4. So, we are trying to say that ρn is sum over m ρm πmn. So, what this means is at the stationary solution for example for n equal to 1, $\rho_1 = \rho_1 \pi_{11} + \rho_2 \pi_{21} + \rho_3 \pi_{31} + \rho_4 \pi_{41}$.

It is some kind of a material balance that we can think of as chemical engineers. So, if we are here either we were here in the last step or we came from here or we came from here or we came from here. So, in the new state the probability density is given by if I have been here multiplied by the transition probability of remaining wherever I was,

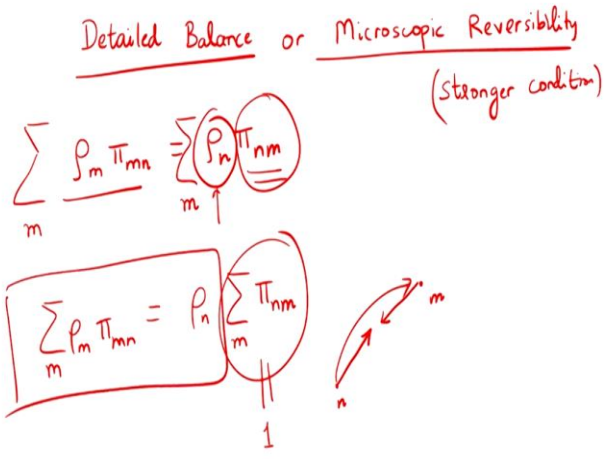
$$\rho_1 = \rho_1\pi_{11} + \rho_2\pi_{21} + \rho_3\pi_{31} + \rho_4\pi_{41}$$

(Refer Slide Time: 27:29)



So, although this may seem quite natural to be valid we have to keep in mind is that this is like true for some sort of a stationary solution. So, these quantities are no longer changing over time, so this should be like valid when the rho's kind of converge to a fixed value at a given phase point, over the course of the Monte Carlo simulations the rho's may change. But we have to wait until that stationary solution is obtained and only for that this is this particular relation is kind of valid. So, there is one more way to put this relation this is called somewhat a weaker condition, but there is a stronger condition that also ensures this particular weak condition and that condition is referred as the condition of microscopic reversibility or detailed balance.

(Refer Slide Time: 31:25)



So, in detailed balance condition or the condition of Microscopic Reversibility, what we say is

$$\rho_m \pi_{mn} = \rho_n \pi_{nm}$$

and that is essentially means that if I look at two states in the state space, let us say state n and state m. Then the ρ multiplied by π that is how much of n is going to m should be matched by how much of m goes to n and that can be thought of as a condition of equilibrium. So, the forward and reverse processes are having the same rate that is the condition of equilibrium.

So, the number of ways in order to speak in which I am going from state n to state m is kind of compensated by the number of ways I am coming from m to n and what has to be kept in mind is that I am actually representing the number of ways by the probability density multiplied with the transition probabilities and we have some control on the choice of the transition probabilities and we have I would say some control on ρ as well, but we have to wait until the stationary solution when the ρ become constant. So, we still have some control over how I am making the transition and no matter what is the function that I have working with whatever π function I pick it should be such that we should always go towards the condition of detailed balance as long as the detailed balance is satisfied that is called a stronger condition we can rest ensured that we will get to the equilibrium or stationary solution and therefore the Monte Carlo simulation scheme is valid.

So, we can have any choice of π because in any case we are not bothered about the actual dynamics. So, what when we say that I jump from state n to state m that is not really happening in any actual system that is simply our way to sample the phase space that we are allowing transition from state n to state m there is nothing like an actual process going from here to there, it is only happening on a computer just to sample the phase space whatever we are doing is simply a numerical scheme so as to speak to sample the phase space there is no dynamics in there.

So, I have full control on π matrix that I pick provided that I have made a choice that satisfies the detailed balance so most algorithms therefore satisfy the detailed balance condition it turns out that even if the detailed balance the condition that I wrote is not valid, but this weaker condition is valid even then you can build some Monte Carlo simulation schemes which are still giving me the equilibrium behavior, but for most schemes we look for the detailed balance condition because that ensures the weaker condition we can see why it is valid.

So, let us say for example if I sum both sides by m here, so on the right hand side ρn does not depend on n. So, I can move that on the other side, so we have-

$$\sum_m \rho_m \pi_{mn} = \rho_n \sum_n \pi_{nm}$$

and what is that so I am basically adding the transition probability of starting from state n and going to every other possible state. So, m is a sum over every possible state, so π_{nm} therefore must sum to 1, because from state n I can only go to the available state space. So, if I sum over all the probabilities from starting from state n that must should go that must go to 1 and that is the condition that was the weaker condition I discussed earlier.

So, using detailed balance I can get the weaker condition, but using the weaker condition I cannot get the detailed balance, detailed balance I would say is therefore is a more stronger condition and as long as the detailed balance is valid we can pretty much choose any transition probability of interest whatever helps me sample the phase space in an efficient manner. So, with that I want to conclude the discussion today in the next lecture we will discuss one type of the choice of π or one type of the Monte Carlo scheme that is known as the metropolis algorithm and I will also touch upon the idea of boundary conditions.

So, with that I conclude here thank you.