Chapter 9

Monte Carlo Simulations

So far we have either dealt with exactly soluble systems like the (time-dependent) oscillator or with various expansions like the semiclassical, perturbative and high-temperature expansion. In this section I introduce a numerical method which plays an important role in modern developments in (gauge-) field theories - the *Monte-Carlo simulations*.

One conveniently starts from the lattice approximation to $Z(\beta, q) = K(\hbar\beta, q, q)$ in (6.21) for the density of the partition function. In the finite dimensional integral over the points $w_1, \ldots, w_{n-1}$ defining the broken line path one sets $w_0 = w_n = q$ and integrates over q to get the trace $Z(\beta)$. More generally, the lattice approximation of expectation values reads

$$\langle A \rangle = \frac{\int d^n w \, A(w) \, e^{-S_E(w)}}{\int d^n w e^{-S_E(w)}}, \quad \text{where} \quad \int d^n w = \int_{-\infty}^{\infty} \prod_1^n dw_j, \tag{9.1}$$

and $S_E(w) = S_E(w_1, \ldots, w_n)$ is the Euclidean lattice-action entering the discrete path integral (6.21). One could use standard numerical integration routines and approximate an integral by a sum with a finite number of terms

$$\int d^n w f(w) \sim \sum_{\mu=1}^{M} f(w_\mu) \Delta w_\mu.$$

Note that w_μ denotes a configuration labelled by μ whereas w_j denotes the j'th component of the configuration $w = (w_1, \ldots, w_n)$. the If the space of integration is high-dimensional, it is however advantageous to choose the points w_μ at random instead of using a regular set of points. Note, however, that the integrand $\exp(-S)$ will vary over many orders of magnitude for different configurations w_μ (we used that already in the semiclassical approximation). The Monte Carlo method introduced by METROPOLIS, ROSENBLUTH, ROSENBLUTH and TELLER is based on the idea of *important sampling* [29]. Instead of selecting the points w_μ at random they are taken from that region where the dominant contributions to the integrals come from. More precisely, we choose the points according to the Boltzmann distribution

$$P(w) = \frac{1}{Z} e^{-S_E(w)}, \quad Z = \int e^{-S_E(w)} \, d^n w \tag{9.2}$$

such that the Monte Carlo estimate $\bar{A}$ for the average of A simply reduces to an arithmetic average

$$\langle A\rangle \sim \bar{A} = \frac{1}{M}\sum_{\mu=1}^{M} A(w_\mu), \tag{9.3}$$

where M is the total number of points (states) generated in the Monte Carlo simulation.

9.1 Markov Processes and Stochastic Matrices

We now discuss a realization of important sampling. It is not quite straightforward, since P is unknown. But it is possible to construct a *Markov process* generating the M configurations w_μ according to given transition probabilities $W(w_\mu \to w_\nu)$ (the probability that the system, currently in configuration w_μ makes a one-step transition into the configuration w_ν). Being a transition probability, the stochastic matrix W must be positive and normalized.

To proceed we consider a system with a finite number f of degrees of freedom. We denote the states by $s \in 1, 2, \ldots, f$ and the transition probabilities by $W(s \to s') \equiv W(s, s')$. They are positive,

$$W(s, s') \geq 0 \tag{9.4}$$

and normalized

$$\sum_{s'} W(s, s') = 1. \tag{9.5}$$

A f-dimensional matrix W with these properties is called *stochastic matrix*. In a two-step process from s to s' the system must pass through some intermediate state s_1 such that the probability of a transition from s to s' in two steps is given by

$$W^{(2)}(s, s') = \sum_{s_1} W(s, s_1)W(s_1, s'). \tag{9.6}$$

Similarly for an n-step process, we have

$$W^n(s, s') = \sum_{s_1 \cdots s_{n-1}} W(s, s_1)W(s_1, s_2)\cdots W(s_{n-1}, s') = \sum_{s_1} W^{n-1}(s, s_1)W(s_1, s'). \tag{9.7}$$

Now one tries to construct a Markov process such that in the 'long-time' limit $n \to \infty$ the configurations are distributed according to (9.2). The 'long-time' behavior of the system is determined by the limit of W^n as $n \to \infty$.

The set of stochastic matrices form a semigroup and every stochastic matrix transforms a stochastic vector (a f-dimensional vector $\boldsymbol{p}$ with non-negative entries P_s obeying $\sum P_s = 1$) into a stochastic vector. The entry P_s of $\boldsymbol{p}$ is the probability to find the system in the state s.

Consider for example a system with 2 states and stochastic matrix

$$W = \begin{pmatrix} a & 1-a \\ 0 & 1 \end{pmatrix} \quad \text{with} \quad W^n = \begin{pmatrix} a^n & 1-a^n \\ 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \tag{9.8}$$

for $a < 1$. The stochastic matrices W^n converges exponentially fast as $n \to \infty$. Also note that this matrix has the eigenvalue 1. Later we shall see that under certain assumptions on W the W^n always converge to a stochastic matrix with identical rows.

A second and less trivial example is the stochastic matrix for a system with 3 states,

$$W = \begin{pmatrix} a & \frac{1}{2}(1-a) & \frac{1}{2}(1-a) \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \tag{9.9}$$

the powers of which are

$$W^n = \begin{pmatrix} a^n & \frac{1}{2}(1-a^n) & \frac{1}{2}(1-a^n) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{for even} \quad n$$

and

$$W^n = \begin{pmatrix} a^n & \frac{1}{2}(1-a^n) & \frac{1}{2}(1-a^n) \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \qquad \text{for odd} \quad n,$$

and which again possesses the eigenvalue 1. For $a < 1$ a stochastic vector $\boldsymbol{p}$ is mapped into

$$\boldsymbol{p} \longrightarrow \boldsymbol{p}\, W^n \longrightarrow \left(0, p_2 + \frac{p_1}{2}, p_3 + \frac{p_1}{2}\right) \quad \text{or} \quad \left(0, p_3 + \frac{p_1}{2}, p_2 + \frac{p_1}{2}\right)$$

as $n \to \infty$, depending on whether n is even or odd. Note that $\boldsymbol{p}\, W^n$ approaches a periodic orbit exponentially fast and hence the Markov process does not converge. As we shall see, the reason for this lack of convergence is that the minima in all columns of W are zero.

A stochastic matrix has always the eigenvalue 1. The corresponding right-eigenvector is $(1, 1, \ldots, 1)^T$. To find the left-eigenvector we consider the sequence

$$\boldsymbol{p}_n = \frac{1}{n} \sum_{j=0}^{n-1} \boldsymbol{p}\, W^j. \tag{9.10}$$

Since the set of stochastic vectors is compact this series must have a convergent subsequence

$$\frac{1}{n_k} \sum_{0}^{n_k - 1} \boldsymbol{p}\, W^j \longrightarrow \boldsymbol{P}.$$

We multiply with W from right and find

$$\frac{1}{n_k}\sum_1^{n_k} \boldsymbol{p}\, W^j \longrightarrow \boldsymbol{P}W.$$

Subtracting the two series yields

$$\frac{1}{n_k}\left(\boldsymbol{p} - \boldsymbol{p}\, W^{n_k}\right) \longrightarrow \boldsymbol{P} - \boldsymbol{P}W$$

For $n_k \to \infty$ the left hand side tends to zero and hence

$$\boldsymbol{P}W = \boldsymbol{P}. \tag{9.11}$$

and therefore any stochastic matrix W possesses at least one fixpoint $\boldsymbol{P}$ that is a eigenvector with eigenvalue 1. Let us now assume that W possesses at least one column whose entries are bigger or equal to a positive number δ which means that all states have a non-vanishing probability for a transition in a certain state. Such W are called *attractive stochastic matrices*. The W in the first example above is attractive whereas the second one is not. For an attractive W all states s have a non-vanishing probability to end up in a given state s', that is $W(s, s') > 0$ for all s. Now we prove that an attractive W is contractive on vectors of length 2. First we note that for two real numbers p and p' we have

$$|p - p'| = p + p' - 2\min(p, p'),$$

such that for two stochastic vectors

$$\|\boldsymbol{p} - \boldsymbol{p}'\| = 2 - 2\sum_s \min(p_s, p'_s). \tag{9.12}$$

Now we prove that an attractive W is contractive on vectors $\boldsymbol{\Delta} = (\triangle_1, \ldots, \triangle_f)$ with

$$\|\boldsymbol{\Delta}\| \equiv \sum |\triangle_s| = 2 \quad \text{and} \quad \sum \triangle_s = 0. \tag{9.13}$$

First we prove this statement for the difference of two cartesian basis vectors $\boldsymbol{e}_s,\ s = 1, \ldots, f$. For that we apply the identity (9.12) to the stochastic vectors $\boldsymbol{e}_s W$ and $\boldsymbol{e}_{s'} W$, that is to the rows s and s' of W . For an attractive W we find for $s \neq s'$

$$\begin{aligned} \|\boldsymbol{e}_s W - \boldsymbol{e}_{s'} W\| &= 2 - 2\sum_{s''} \min\left\{W(s, s''), W(s', s'')\right\} \\ &\leq 2 - 2\delta = (1 - \delta)\underbrace{\|\boldsymbol{e}_s - \boldsymbol{e}_{s'}\|}_{=2} \quad \text{with} \quad 0 < \delta < 1, \end{aligned} \tag{9.14}$$

and this proves that W is contractive of the difference vectors $\boldsymbol{e}_s - \boldsymbol{e}_{s'}$. We used

$$\min_{s''}\left\{W(s, s'')W(s', s'')\right\} \geq \min\left\{W(s, s^*)(W(s', s^*)\right\} \geq \delta$$

where s^* belongs to the particular column of W the elements of which are bigger or equal to δ.

Now we prove the contraction property for all vectors $\boldsymbol{\Delta}$ in (9.13). Since

$$\begin{aligned}\sum_{s:\triangle_s\geq 0}\triangle_s-\sum_{s:\triangle_s<0}\triangle_s &= \|\boldsymbol{\Delta}\|=2\\ \sum_{s:\triangle_s\geq 0}\triangle_s+\sum_{s:\triangle_s<0}\triangle_s &= 0\end{aligned}$$

we conclude that

$$\sum_{\triangle_s\geq 0}\triangle_s=1 \quad\text{and}\quad \sum_{\triangle_s<0}\triangle_s=-1. \tag{9.15}$$

To simplify our notation we denote the non-negative elements of $\boldsymbol{\Delta}$ by $\triangle_s$ and the negative elements by $\triangle_{s'}$. Note that the index sets $\{s\}$ and $\{s'\}$ have no common elements. Because of (9.15) we have

$$\|\boldsymbol{\Delta}\|=2=-2\sum\triangle_s\sum\triangle_{s'}=-\sum\triangle_s\triangle_{s'}\underbrace{\|\boldsymbol{e}_s-\boldsymbol{e}_{s'}\|}_{=2}, \tag{9.16}$$

where we assumed $s\neq s'$. To bound the norm of $\boldsymbol{\Delta}W$ we use

$$\sum\triangle_s\boldsymbol{e}_s=-\sum\triangle_{s'}\sum\triangle_s\boldsymbol{e}_s \quad,\quad \sum\triangle_{s'}\boldsymbol{e}_{s'}=+\sum\triangle_s\sum\triangle_{s'}\boldsymbol{e}_{s'},$$

where we made use of (9.15), and this leads to the inequality

$$\begin{aligned}\|\boldsymbol{\Delta}W\| &= \|\sum\triangle_s\boldsymbol{e}_sW+\sum\triangle_{s'}\boldsymbol{e}_{s'}W\|=\|-\sum\triangle_{s'}\triangle_s(\boldsymbol{e}_s-\boldsymbol{e}_{s'})W\|\\ &\leq-\sum\triangle_s\triangle_{s'}\|(\boldsymbol{e}_s-\boldsymbol{e}_{s'})W\|\leq-\sum\triangle_s\triangle_{s'}\|\boldsymbol{e}_s-\boldsymbol{e}_{s'}\|(1-\delta),\end{aligned} \tag{9.17}$$

where we used the inequality (9.14). Comparing with (9.16) leads to the inequality

$$\|\boldsymbol{\Delta}\,W\|\leq(1-\delta)\|\boldsymbol{\Delta}\| \tag{9.18}$$

which shows that W is contractive on vectors of the form (9.13). Since this inequality is linear $\boldsymbol{\Delta}$ we may drop the condition $\|\boldsymbol{\Delta}\|=2$. Hence W is contractive on all vectors the elements of which add up to zero and in particular on differences of two stochastic vectors.

Iterating the inequality we obtain

$$\|\boldsymbol{\Delta}\,W^n\|\leq(1-\delta)^n\|\boldsymbol{\Delta}\,\|. \tag{9.19}$$

Now we apply this inequality to $\boldsymbol{p}-\boldsymbol{P}$, where $\boldsymbol{P}$ is the fixpoint in (9.11) and $\boldsymbol{p}$ is an arbitrary stochastic vector. Since the elements of $\boldsymbol{p}-\boldsymbol{P}$ add up to zero we conclude

$$\|(\boldsymbol{p}-\boldsymbol{P})W^n\|=\|\boldsymbol{p}\,W^n-\boldsymbol{P}\|\overset{n\to\infty}{\longrightarrow}0$$

or equivalently that

$$\boldsymbol{p}\, W^n \overset{n\to\infty}{\longrightarrow} \boldsymbol{P}. \tag{9.20}$$

For the stochastic vectors $\boldsymbol{e}_s$ the left hand side is just the row s of the limit $\lim W^n = W^{\rm eq}$ such that $W^{\rm eq}$ has identical rows. This means that all elements in a column of $W^{\rm eq}$ are equal, similarly as in the example above.

$$W^{\rm eq}(s,s') = \lim_{n\to\infty} W^n(s,s') = P_{s'}, \tag{9.21}$$

where $P_{s'}$ is the element s' of $\boldsymbol{P}$. It follows that both the limit distribution $\boldsymbol{P}$ and matrix $W^{\rm eq}$ *are unique*. Else there would exist a second fixpoint $\boldsymbol{P'}$ of the Markov process with

$$P'_{s'} = \sum_s P'_s W(s,s') = \lim_{n\to\infty} \sum_s P'_s W^n(s,s') = \sum_s P'_s P_{s'} = P_{s'},$$

and this shows that $\boldsymbol{P}$ is indeed the unique fixpoint.

The generalization to systems with continuous degrees of freedom is clear. For the discretized path integral the states of a Markov process are the broken line paths $w = (w_1, \ldots, w_n)$ on a time lattice with n sites. The probability P_s becomes a *probability density* $P(w)$. Sums over the discrete index s become integrals over the continuous variables $w_1, \ldots, w_n$. For the discretized path integral the conditions (9.4) and (9.5) read

$$W(w,w') \geq 0 \quad \text{and} \quad \int \mathcal{D}w'\, W(w,w') = 1. \tag{9.22}$$

The fixpoint condition for the equilibrium distribution $P(w)$ takes the form

$$P(w') = \int \mathcal{D}w\, P(w)\, W(w,w'). \tag{9.23}$$

In Euclidean Quantum Mechanics or in Quantum Statistics the equilibrium distribution P is the the Boltzmann distribution, see (9.2).

9.2 Detailed Balance, Metropolis Algorithm

The interesting question is whether we can find a simple stochastic process which has the Boltzmann distribution as its fixpoint. All processes used in actual simulations are constructed with stochastic matrices fulfilling the *detailed balance* condition

$$P_s W(s,s') = P_{s'} W(s',s), \tag{9.24}$$

which means that the transition from s to s' is more probable as the inverse process if the equilibrium density is bigger at s' than it is at s. If the detailed balance condition holds then $\boldsymbol{P}$ is indeed a fix point of W since

$$\sum_s P_s W(s,s') = \sum_s P_{s'} W(s',s) = P_{s'}. \tag{9.25}$$

The detailed balance condition does not fix the stochastic process uniquely. Considerations of computational efficiency dictates its specific form. The local *Metropolis-* and *heatbath algorithm* are the most popular once since they can be applied to almost all statistical systems. More recently nonlocal *cluster algorithms* are also used in cases where local algorithms fail due to 'critical slowing down'. Unfortunately cluster algorithms are not known for lattice gauge theories.

In the METROPOLIS algorithm one starts with a certain initial configuration. The better the initial configuration is chosen the less computer time is needed. For example, for a statistical lattice model at high temperature one should choose the variables at different lattice sites (having the lattice approximation (6.21) in mind) randomly, since there is almost no correlation between variables at neighboring sites (see the high temperature expansion). If we are simulating a quantum mechanical system with a deep double well potential then (at low or zero temperature) one better chooses a initial configuration which resembles an instanton. After preparing the initial configuration one takes the 'first' lattice variable w_1 and changes it or leaves it according to the following rules:

- First one replaces w_1 tentatively by a *randomly* chosen trial-variable w_1'.
- If S decreases (P increases) then the variable at site 1 is set to the new value w_1'.
- If the action increases then a random number r with uniform distribution between 0 and 1 is generated and the variable is changed to w_1' only if $\exp(-\Delta S) > r$. Otherwise the lattice variable retains its previous value w_1.
- The other sites of the lattice are then probed in exactly the same way.
- When one arrives at the last lattice site then one has completed one *Monte-Carlo sweep*.

To test whether one is already near equilibrium one measures some expectation values as a function of iteration time. After the measured values have settled (that is we are not too far from the equilibrium distribution) then we may start 'measuring' observables according to (9.3).

The Heat-bath algorithm is very similar as the Metropolis algorithm and I refer you to the literature on Monte Carlo simulations for learning more about the various techniques to simulate quantum- or spin system. For an introduction into Monte Carlo (MC) methods, numerical studies and further references I refer you to the nice paper of Creutz and Freedman [30], the text book of NEWMAN and BARKENNA [31] or the text book of BERG [32].

9.2.1 Three-state system at finite temperature

Let us finally consider a physical system with 3 states.

$$H|n\rangle = E_n|n\rangle\,, \quad n = 1, 2, 3 \quad \text{with} \quad E_1 < E_2 < E_3. \tag{9.26}$$

Since the energy decreases for the transitions $|2\rangle \to |1\rangle$, $|3\rangle \to |1\rangle$ and $|3\rangle \to |2\rangle$ the corresponding transition amplitudes are constant in the Metropolis algorithm. The inverse transitions cost energy and the probabilities are proportional to the Boltzmann factors

$$b_{pq} = e^{\beta(E_p - E_q)}, \quad p > q. \tag{9.27}$$

We see that the Markov process has the stochastic matrix

$$W = \frac{1}{2}\begin{pmatrix} 2 - b_{21} - b_{31} & b_{21} & b_{31} \\ 1 & 1 - b_{32} & b_{32} \\ 1 & 1 & 0 \end{pmatrix}, \tag{9.28}$$

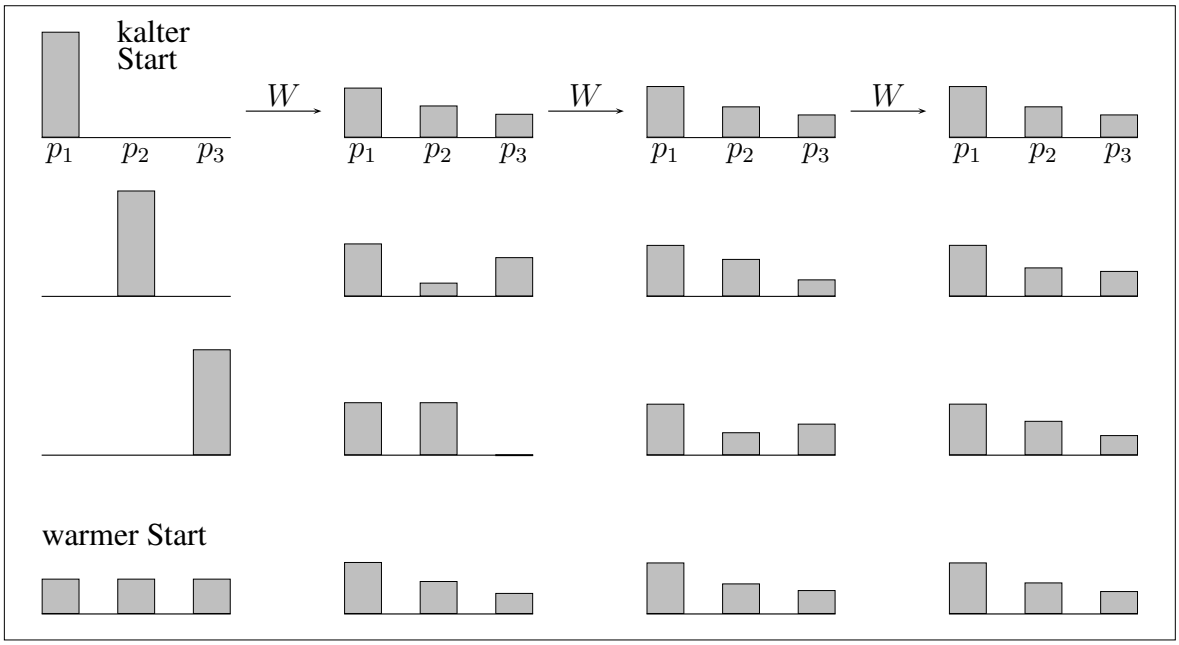
and its powers converge to

$$W^\infty = \frac{1}{Z}\begin{pmatrix} e^{-\beta E_1} & e^{-\beta E_2} & e^{-\beta E_3} \\ e^{-\beta E_1} & e^{-\beta E_2} & e^{-\beta E_3} \\ e^{-\beta E_1} & e^{-\beta E_2} & e^{-\beta E_3} \end{pmatrix}, \tag{9.29}$$

where Z is the partition function, $Z = \exp(-\beta E_1) + \exp(-\beta E_2) + \exp(-\beta E_3)$. Every initial probability distribution converges to the Boltzmann distribution,

$$\boldsymbol{P} = \frac{1}{Z}\left(e^{-\beta E_1}, e^{-\beta E_2}, e^{-\beta E_3}\right). \tag{9.30}$$

The following figure shows the convergence to equilibrium for different initial distributions. I used the differences $E_2 - E_1 = 0.5$ and $E_3 - E_2 = 0.3$.



For a cold start with ground state as initial state and a hot start with equal probabilities for the different states we find good convergence to equilibrium. If we start with the most excited state as initial state the the convergence is bad.