Quantum Monte Carlo techniques

We have previously seen that we can solve a quantum lattice model of interest via exact diagonalization, thereby obtaining the lowest eigenvalues and eigenvector of the Hamiltonian matrix. However, there is a huge cost associated with obtaining exact results – we are limited to rather modest system sizes.

The next step which we will take is to trade the exact solution of a quantum lattice model with a controlled approximation that will allow to converge towards the exact solution with arbitrary precision. We have seen this before – statistical approaches, in particular Monte Carlo approaches, can do precisely this. So we want to set out and learn about such Monte Carlo approaches to quantum (lattice) models – so called quantum Monte Carlo.

In a quantum Monte Carlo simulation we want to evaluate thermodynamic averages such as

$$\langle A \rangle = \frac{\mathrm{Tr}\ A \cdot e^{-\beta H}}{\mathrm{Tr}\ e^{-\beta H}}$$

The main problem is the calculation of the exponential $e^{-\beta H}$. A straightforward calculation would require a complete diagonalization which is just what we want to avoid.

The essential idea is to find a mapping of the quantum partition function to a classical problem

$$Z = \mathrm{Tr}\, e^{-\beta H} \equiv \sum_c p_c$$

with the additional constraint that $p_c > 0$ for all configuration of the classical problem — otherwise we would face the so-called "negative sign" problem if some $p_c < 0$.

The idea for the concrete mapping goes back to Richard Feynman (1953): a d-dimensional quantum system can be mapped to a $(d+1)$-dimensional classical system using a path integral representation. It is precisely this idea on which all quantum Monte Carlo algorithms are based.

Path integral representation in terms of world lines

Let's see how we can perform such a mapping taking a concrete example — the spin-½ quantum XXZ model with Hamiltonian

$$H_{XXZ} = -\sum_{\langle ij \rangle} \left(J_z S_i^z S_j^z + J_{xy} \left(S_i^x S_j^x + S_i^y S_j^y \right) \right)$$

$$= -\sum_{\langle ij \rangle} \left(J_z S_i^z S_j^z + \frac{J_{xy}}{2} \left(S_i^+ S_j^- + S_i^- S_j^+ \right) \right)$$

For $J \equiv J_z = J_{xy}$ we have the Heisenberg model ($J>0$: FM, $J<0$: AFM). $J_{xy}=0$ is the (classical) Ising model and $J_z=0$ is the quantum XY model

For this concrete problem we want to address how we can deal with the calculation of the exponential $e^{-\beta H}$.

- The first step is to identify this exponential with an imaginary time evolution. Recall that time evolution in quantum mechanics is given by

$$e^{-itH}$$

So we can think of the Boltzmann factor $e^{-\beta H}$ as an evolution in imaginary time

$$e^{-\beta H} = e^{-itH}$$

with $t = -i\beta$, i.e. the inverse temperature β sets the length of this imaginary time evolution.

- The second step in calculating the Boltzmann factor is to discretize the imaginary time (inverse temperature) direction and subdivide $\beta = M \cdot \Delta\tau$ so that we have

$$e^{-\beta H} = \left(e^{-\Delta\tau \cdot H}\right)^M = (1 - \Delta\tau H)^M + O(\Delta\tau)$$

In the limit $M \to \infty$ $(\Delta\tau \to 0)$ this becomes exact. We will take the limit later, but stay at finite $\Delta\tau$ for now.

- The third step is to insert the identity matrix, represented by a sum over all basis states $1 = \sum_i |i\rangle\langle i|$ between all operator $(1 - \Delta\tau H)$:

$$Z = \mathrm{Tr}\, e^{-\beta H} = \mathrm{Tr}\, (1 - \Delta\tau H)^M + O(\Delta\tau)$$

$$Z = \sum_{i_1, i_2, \ldots, i_M} \langle i_1 | 1 - \Delta\tau H | i_2 \rangle \langle i_2 | 1 - \Delta\tau H | i_3 \rangle \ldots \langle i_M | 1 - \Delta\tau H | i_1 \rangle + O(\Delta\tau^2)$$

$$\equiv P_{i_1, i_2, \ldots, i_M} \qquad (*)$$

and similarly for the measurement, obtaining

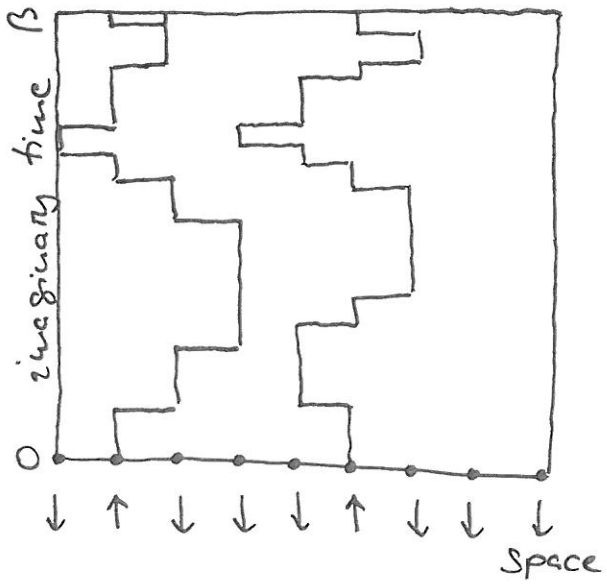
$$\langle A \rangle = \sum_{i_1, \ldots, i_M} \frac{\langle i_1 | A (1 - \Delta\tau H) | i_2 \rangle}{\langle i_1 | 1 - \Delta\tau H | i_2 \rangle} P_{i_1, \ldots, i_M} + O(\Delta\tau^2).$$

If we choose the basis states $|i\rangle$ to be eigenstates of the local S^z operators we end up with an Ising-like spin system in one higher dimension.

Each choice $i_1, \ldots, i_M$ corresponds to one of the possible configurations of this classical spin system.

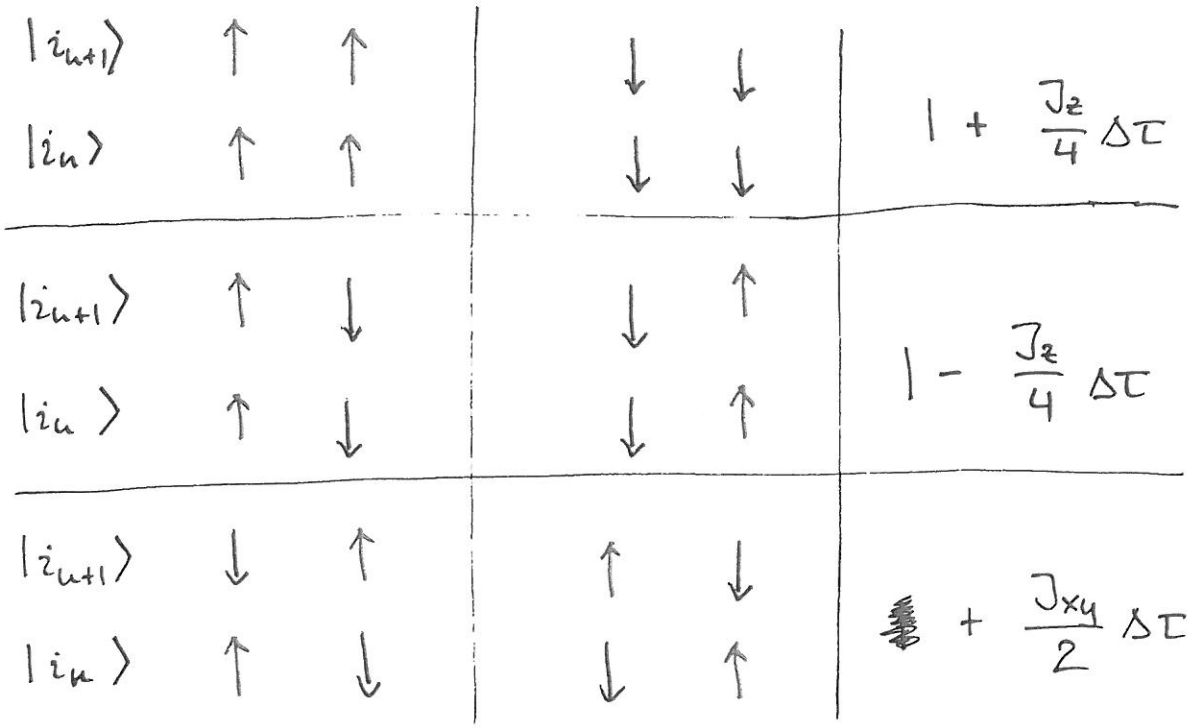
The trace is mapped to periodic boundary conditions in the imaginary time direction of this classical spin system.

Here is an example for a "world line" configuration of this spin-½ model.
Drawn are the world lines for spin-up only. Down-spin world lines occupy the rest of the configuration.



- We can now sample this classical system using classical Monte Carlo methods for which the probabilities are given by matrix elements $\langle i_n | 1 - \Delta\tau H | i_{n+1} \rangle$.

However, most of the matrix elements $\langle i_n | 1 - \Delta\tau H | i_{n+1} \rangle$ are zero, and thus nearly all configurations have vanishing weight. The only non-zero configurations are those where neighboring states $|i_n\rangle$ and $|i_{n+1}\rangle$ are either equal or differ by one of the off-diagonal matix elements in H, which are nearest neighbor exchanges by two opposite spins.

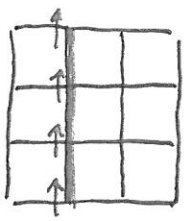


We can thus uniquely connect spins on neighboring "time slices" and end up with world lines of spins.
Instead of sampling over all configurations of local spins we thus have to sample only over all world line configurations. Update moves are not allowed to break world lines but have to lead to new valid world line configurations

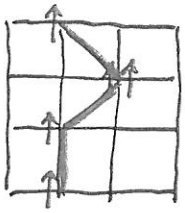
World lines - local updates

Let's take a closer look at the weight of certain world line configurations and construct local update moves that change these world line configurations in a consistent way (i.e. yielding a new valid world line configuration).
For simplicity, we will do this for the XY limit of our spin Hamiltonian i.e. $J_z = 0$.

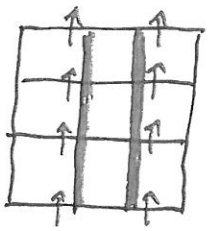
Example world line configurations and their weights:



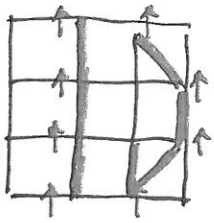
1



$\left(\frac{J_{xy}}{2}\Delta\tau\right)^2$



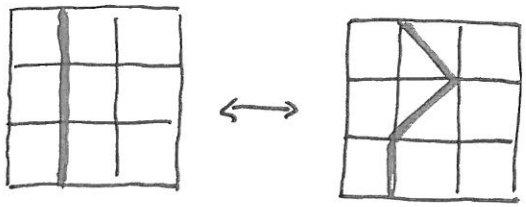
1



$\left(\frac{J_{xy}}{2}\Delta\tau\right)^2$

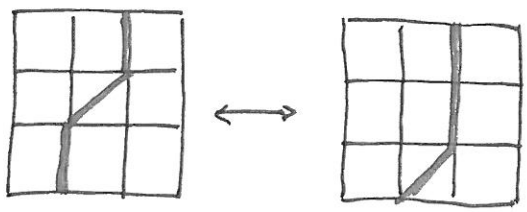
Local updates:

"insert or remove two kinks"



1 $\qquad$ $\left(\frac{J_{xy}}{2}\Delta\tau\right)^2$

"shift a kink"



$\frac{J_{xy}}{2}\Delta\tau$ $\qquad$ $\frac{J_{xy}}{2}\Delta\tau$

$$P_{\rightarrow} = \min\left(1, \left(\frac{J_{xy}}{2}\Delta\tau\right)^2\right)$$

$$P_{\leftarrow} = \min\left(1, \frac{1}{\left(\frac{J_{xy}}{2}\Delta\tau\right)^2}\right)$$

$$P_{\rightarrow} = P_{\leftarrow} = 1$$

The continuous time limit

The systematic error arising from the finite time step $\Delta\tau$ was originally controlled by an extrapolation to the continuous time limit $\Delta\tau \to 0$ from simulations with different values of the time step $\Delta\tau$.

It was later realized that for a discrete quantum lattice model, this limit can already be taken during the construction of the algorithm and simulations can be performed directly at $\Delta\tau \to 0$, corresponding to an infinite number of Trotter steps $M = \infty$.

In this limit the Suzuki-Trotter formula (*) becomes equivalent to a time-dependent perturbation theory in imaginary time.

$$Z_1 = \mathrm{Tr}\, \exp(-\beta \mathcal{H}) = \mathrm{Tr}\left[\exp(-\beta H_0)\, \mathcal{T} \exp \int_0^\beta d\tau\, V(\tau) \right]$$

(time-ordering)

where we have split the original Hamiltonian $\mathcal{H} = H_0 + V$ into diagonal (H_0) and off-diagonal (V) parts. The time-dependent perturbation $V(\tau)$ in the interaction representation is

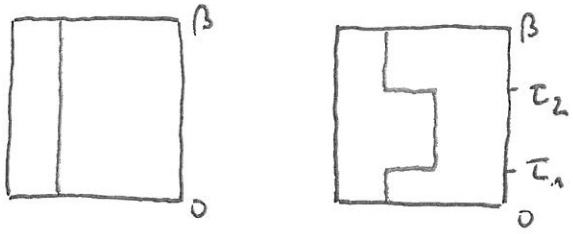
$$V(\tau) = \exp(\tau H_0)\, V \exp(-\tau H_0) .$$

Expanding the time ordering $\mathcal{T}$ of the exponential we obtain

$$Z_1 = \mathrm{Tr}\left[\exp(-\beta H_0) \left(1 - \int_0^\beta d\tau\, V(\tau) + \frac{1}{2} \int_0^\beta d\tau_1 \int_{\tau_1}^\beta d\tau_2\, V(\tau_1) V(\tau_2) + \dots \right) \right]$$

This again leads to a world-line representation.

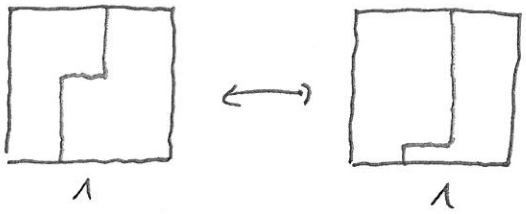
Examples:



1 $\qquad\qquad$ $\left(\frac{J_{xy}}{2}\right)^2 d\tau_1 d\tau_2$

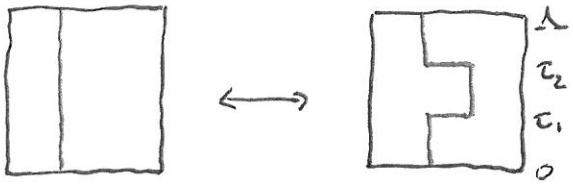
Local updates in continuous time:

- shift a kink to a new position



$P_{\rightarrow} = P_{\leftarrow} = 1$

- insert a pair of kinks



1 $\qquad\qquad$ $\left(\frac{J_{xy}}{2}\Delta\tau\right)^2 \underset{\Delta\tau\to 0}{\longrightarrow} 0$

$$P_{\rightarrow} = \min\left(1, \left(\frac{J_{xy}}{2}\Delta\tau\right)^2\right) \underset{\Delta\tau\to 0}{\longrightarrow} 0$$

solution: integrate over all possible insertions in an interval Λ

$$P = \int_0^{\Lambda}\int_{\tau_1}^{\Lambda} \left(\frac{J_{xy}}{2}\right)^2 d\tau_1 d\tau_2 = \frac{\Lambda^2}{2}\left(\frac{J_{xy}}{2}\right)^2 \neq 0$$

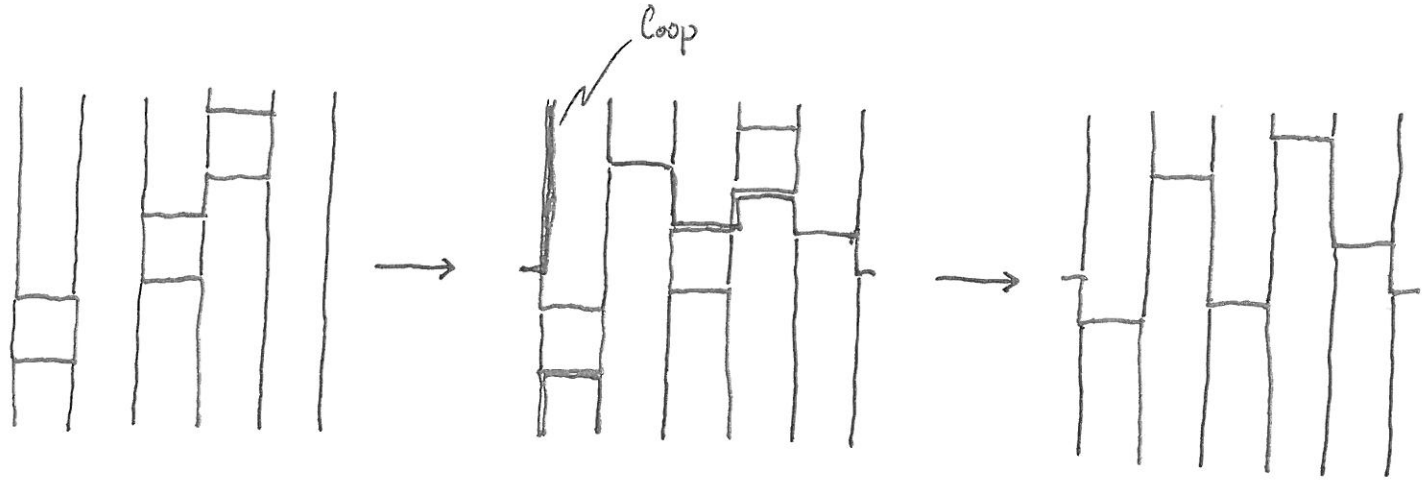
$$P_{\rightarrow} = \min\left(1, \frac{\Lambda^2}{2}\left(\frac{J_{xy}}{2}\right)^2\right) \neq 0.$$

Local versus non-local world line updates

While we have seen that local updates can be used to sample (valid) world line configurations, these updates are intrinsically very limited: they cannot change global or topological properties such as

- the number of world lines (= particles, magnetization) is conserved
- the winding conserved
- the braiding conserved
- cannot sample grand-canonical ensemble

The solution to overcome these limitations are non-local or cluster updates, which in a nutshell can be illustrated via the following loop update:



initial world line configuration

world line conf. plus loop cluster

world line conf. after flipping all spins along the loop

Note the non-trivial winding!

In addition, non-local / cluster updates are again needed to overcome the critical slowing down at second order (thermal) phase transitions when using local updates.

Finally, we note that similar to the classical situation we can ease (or sometimes overcome) the equilibration problem at first-order transitions by introducing an extended ensemble, e.g. parallel tempering or a generalized form of broad histogram sampling.

Overview of current limits for QMC

The scaling of the non-local update QMC techniques is $O(N\beta)$. This polynomial scaling allows to study quite large systems:

temperature	local Metropolis	non-local updates
3D T_c	16'000 spins	16'000'000 spins
0.1 J	200 spins	1'000'000 spins
0.005 J	—	50'000 spins
3D T_c	32 bosons	1'000'000 bosons
0.1 t	32 bosons	10'000 bosons.

The stochastic series expansion (SSE)

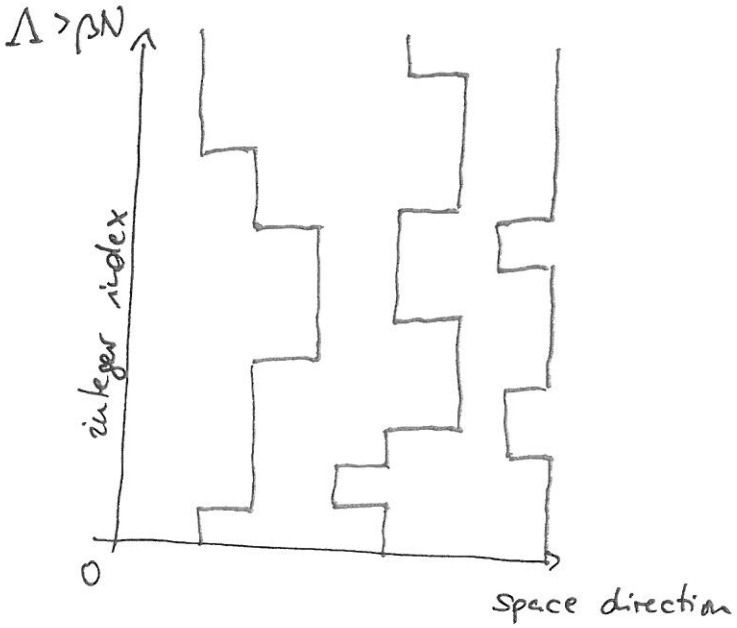
An alternative Monte Carlo algorithm, which also does not suffer from time discretization, is the stochastic series expansion (SSE) algorithm. It starts from a Taylor expansion of the partition function in orders of β:

$$Z = \mathrm{Tr}\, \exp(-\beta \mathcal{H}) = \sum_{n=0}^{\infty} \frac{\beta^n}{n!} \mathrm{Tr}(-\mathcal{H})^n$$

$$= \sum_{n=0}^{\infty} \sum_{i_1, i_2, i_3 \ldots i_n} \frac{\beta^n}{n!} \langle i_1 | -\mathcal{H} | i_2 \rangle \langle i_2 | -\mathcal{H} | i_3 \rangle \ldots \langle i_n | -\mathcal{H} | i_1 \rangle$$

$$= \sum_{n=0}^{\infty} \frac{\beta^n}{n!} \sum_{\{i\}} \sum_{\{b\}} \langle i_1 | -\mathcal{H}_{b_1} | i_2 \rangle \langle i_2 | -\mathcal{H}_{b_2} | i_3 \rangle \ldots \langle i_3 | -\mathcal{H}_{b_n} | i_1 \rangle$$

where we have decomposed the Hamiltonian $\mathcal{H}$ into a sum of single-bond terms $\mathcal{H} = \sum_b \mathcal{H}_b$. We end up with a similar world line representation as in the path integral approach:



The negative sign problem

In the last two lectures we have seen that world-line QMC simulations can be an extremely powerful tool to study quantum lattice models – it is a completely unbiased approach, yet it scales polynomially like $O(\beta N)$ when using non-local update techniques.
If this were the end of the story, we would have practically solved all quantum lattice models. Unfortunately, this is not the case since for some of the most interesting models the QMC approach fails dramatically due to the so-called "negative sign problem". This is what we will talk about today.

Let's recall that at the heart of the QMC approach was a mapping from the quantum mechanical problem to a (classical) system that allowed sampling, i.e

$$Z = \mathrm{Tr}\, e^{-\beta H} \equiv \sum_c p_c$$

where $\{c\}$ is some configuration space, e.g. a $(d+1)$ dimensional system which we obtain via the path-integral world-line representation. So far, we have assumed that a mapping can be found to allow all probabilities p_c to be positive, i.e. $p_c \geq 0 \;\forall c$. What are we supposed to do if some of the p_c turn out to be negative?

Let's see under which circumstances, we can encounter such negative probabilities.

- First, we note that the diagonal part of the Hamiltonian H_0 can always be made positive by a finite off-set (for a given finite system) and thus all diagonal operators can be forced to give positive probabilities.

- The difficult problem thus lies with the off-diagonal terms, e.g. the term

$$V = -\frac{J_{xy}}{2}\left(S_i^+ S_j^- + S_i^- S_j^+\right).$$

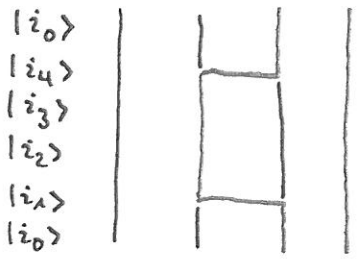
This term has weight $\left(+\frac{J_{xy}}{2}\Delta\tau\right)$ in the Trotter-Suzuki decomposition and similarly positive weight in the SSE representation. So no problem here.

But what about off-diagonal terms with opposite sign

$$V = \oplus\,\frac{J_{xy}}{2}\left(S_i^+ S_j^- + S_i^- S_j^+\right) \quad ?$$

Now we will have to carefully distinguish two cases.

First, let's look at <u>bipartite lattices</u>, i.e. a lattice with an even number of sites per "plaquette".

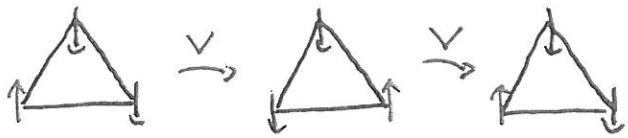


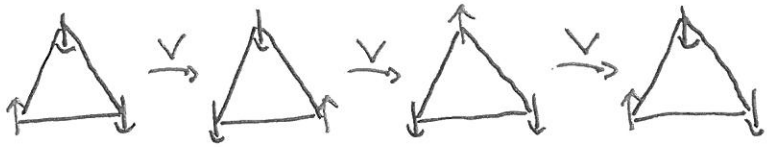
In this case, the off-diagonal operators (or links) must occur in pairs.

Then we always end up with positive weight

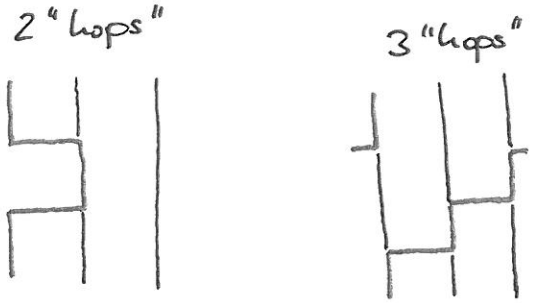
$$\left(-\frac{J_{xy}}{2}\Delta\tau\right)^{2n}$$

Now, let's look at a non-bipartite lattice, i.e. a lattice with an odd number of sites per "plaquette", e.g. a triangular lattice

2 "hops"  $\Rightarrow \left(-\frac{J_{xy}}{2}\Delta\tau\right)^2 > 0$

3 "hops"  $\Rightarrow \left(-\frac{J_{xy}}{2}\Delta\tau\right)^3 < 0$

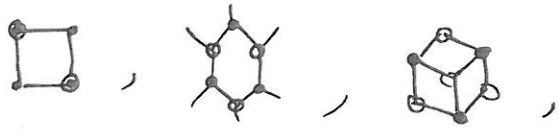
with the corresponding world line configurations



So, we realize that on non-bipartite lattice we can encounter the possibilty of negative weights.

<u>Theorem</u>: If a path can be found through the lattice that takes an odd number of "hops" (i.e. off-diagonal operations of any length), then the sign problem can occur.
(Sometimes a clever basis rotation can get rid of this.)

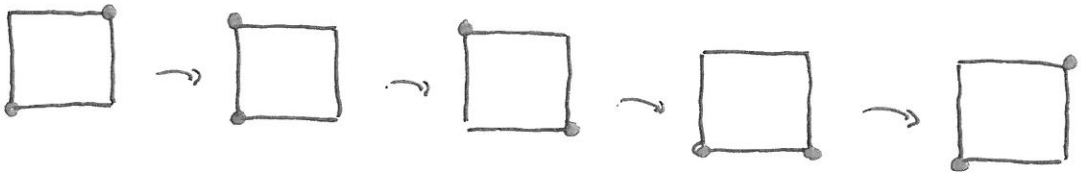
Examples: No sign problem for $H_{XXZ} = \sum_{\langle ij\rangle} \left(S_i^z S_j^z + \frac{1}{2}(S_i^+ S_j^- + S_i^- S_j^+)\right)$

on  , etc.

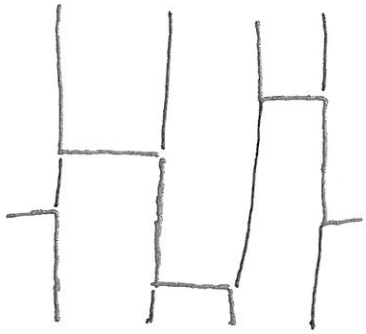
However, this has an uncorrectable sign problem for

pyrochlore

An even more fundamental example for the occurence of the sign problem is found in fermionic systems. Here we will encounter a negative weight for a world-line configuration that includes the (single) exchange of two fermion e.g.



with world-line representation



In fact, the origin of the sign problem in frustrated magnets also arises from the fermionic minus sign of the superexchange process (note that we have swapped the positions of the two down spins in our triangular lattice example).

What can we do when we encounter configurations with negative weights?

One idea is to sample with respect to the absolute values of the weights, i.e.

$$\langle A \rangle = \frac{\sum_c A_c P_c}{\sum_c P_c} = \frac{\sum_c A_c \cdot \mathrm{sgn}\, P_c \cdot |P_c|}{\sum_c \mathrm{sgn}\, P_c \cdot |P_c|}$$

$$= \frac{\sum_c (A_c \cdot \mathrm{sgn}\, P_c) \cdot |P_c| \,/\, \sum_c |P_c|}{\sum_c (\mathrm{sgn}\, P_c) \cdot |P_c| \,/\, \sum_c |P_c|} = \frac{\langle A \cdot \mathrm{sign} \rangle_{|P|}}{\langle \mathrm{sign} \rangle_{|P|}} \quad (*)$$

where we have now included the sign in the observable.

The "sign problem" is the fact that the errors get blown up by an additional factor $1/\langle \mathrm{sign} \rangle_{|P|}$ which grows exponentially with volume and temperature, i.e.

$$\langle \mathrm{sign} \rangle_{|P|} = \frac{\sum_c P_c}{\sum_c |P_c|} = \frac{Z}{Z_{|P|}} = e^{-\beta V (f - f_{|P|})}$$

Z: partition function of fermionic system

$f - f_{|P|}$: difference of free energy densities

$Z_{|P|}$: partition function of bosonic system used to sample

This leads to an exponential growth of statistical errors

$$\frac{\Delta \mathrm{sign}}{\langle \mathrm{sign} \rangle} = \frac{\sqrt{(\langle \mathrm{sign}^2 \rangle - \langle \mathrm{sign}^2 \rangle)/M}}{\langle \mathrm{sign} \rangle} = \frac{\sqrt{1 - \langle \mathrm{sign} \rangle^2}}{\sqrt{M} \cdot \langle \mathrm{sign} \rangle} \sim \frac{e^{+\beta V \Delta f}}{\sqrt{M}}$$

Similarly, the error of the numerator in (*) increases exponentially.

The origin of the sign problem can be tracked back to the simple observation that by ignoring the sign we sample with the wrong distribution.

- We simulate bosons and expect to learn about fermions?
- We simulate a ferromagnet and expect to learn something useful about a frustrated antiferromagnet?
- We simulate a ferromagnet and expect to learn something about a spin glass?

The occurence of the negative weights is of course a basis dependent statement. So, in order to avoid / solve the sign problem we "simply" need to find a basis which does not give rise to negative weights.

However, precisely this challenge to find a suitable basis has been shown to be an NP-hard problem (Troyer and Wiese '05). So there is very little hope that we will find a general solution to this problem. For special situations, we might be lucky, but otherwise we will have to continue our quest of finding a powerful approach to attack the many-fermion problem.

Variational Monte Carlo (VMC)

Quantum Monte Carlo approaches come in many flavors, in particular there are many other methods beyond the world-line Monte Carlo techniques we have talked about in the last two lectures. We chose to take a closer look at those world-line sampling techniques, because they identified a completely unbiased and fully controlled approach to quantum lattice models if they were not rendered inadequate by the (fermion) sign problem.

Before concluding our discussion of QMC techniques, let's therefore take a look at at least one other flavor of QMC – the variational Monte Carlo technique.

As the name suggests VMC builds on the variational method, which we briefly recall in the form of an algorithm:

- Construct a trial many-particle wavefunction $\Psi_\alpha(R)$, which depends on a number of variational parameters $(\alpha_1, \alpha_2, \ldots, \alpha_S)$. The Ψ_α further depends on the combined coordinates / states of N degrees of freedom.
- evaluate the expectation value of the energy

$$\langle E \rangle_\alpha = \frac{\langle \Psi_\alpha | H | \Psi_\alpha \rangle}{\langle \Psi_\alpha | \Psi_\alpha \rangle} \geq E_0$$

- Vary the parameters α according to some minimization algorithm and return to step 1.

We will now play a variant of this variational method where the parameters α are adjusted within the framework of a Monte Carlo simulation. The observation that justifies such an approach is that for realistic systems the many-body wavefunction assumes very small values in large parts of the configuration space and assumes appreciable values only in a small part of the configuration space.

Let's again start with a trial wavefunction $|\Psi_\alpha\rangle$. We would like this trial wave function to be a good representation of the actual ground state of a model. Finding the best (trial) wave function means finding the right set of parameters $\{\alpha\}$ that maximize the overlap with the actual ground state. In practice, this is impossible since we do not know the ground state a priori, and some physical insight is needed to derive a good approximation, i.e. a particular form of the trial wave function.

For a given wave function $|\Psi_\alpha\rangle$ we have already seen that the energy $\langle E\rangle_\alpha$ is always greater or equal to the exact energy of the ground state $\langle E\rangle_\alpha \geq E_0$, which we use to set the criterium that we want to minimize the variational energy.

So, for a given trial wavefunction $|\psi_\alpha\rangle$ we want to calculate $\langle E\rangle_\alpha$. Let's see how we can do this by first considering the expectation value of an arbitrary operator A

$$\langle A\rangle_\alpha = \frac{\langle\psi_\alpha|A|\psi_\alpha\rangle}{\langle\psi_\alpha|\psi_\alpha\rangle}$$

$$= \frac{\sum_x \langle\psi_\alpha|x\rangle\langle x|A|\psi_\alpha\rangle}{\sum_x \langle\psi_\alpha|x\rangle\langle x|\psi_\alpha\rangle}$$

$$= \frac{\sum_x |\langle\psi_\alpha|x\rangle|^2 \cdot \frac{\langle x|A|\psi_\alpha\rangle}{\langle x|\psi_\alpha\rangle}}{\sum_x |\langle\psi_\alpha|x\rangle|^2}$$

$$\equiv \sum_x P_x \cdot A_x \qquad (*)$$

where

$$P_x = \frac{|\langle\psi_\alpha|x\rangle|^2}{\sum_x |\langle\psi_\alpha|x\rangle|^2} \qquad A_x = \frac{\langle x|A|\psi_\alpha\rangle}{\langle x|\psi_\alpha\rangle} \qquad (**)$$

Equation (*) has precisely the form of a mean value in statistical mechanics, with P_x being the statistical weight:

$$P_x \geq 0 \qquad \sum_x P_x = 1$$

In order to extract information (such as the energy) from the trial wavefunction $|\psi_\alpha\rangle$ we will employ this perspective to generate/sample a collection of states according to this distribution.

We know how to do this — we employ a Metropolis algorithm: Starting from a configuration $|x\rangle$, we accept a new configuration $|x'\rangle$ with probability

$$R = \frac{|\langle \psi_\alpha | x' \rangle|^2}{|\langle \psi_\alpha | x \rangle|^2}$$

Since the probabilities do not depend on the statistics of the particles involved, they do not suffer from the sign problem. However, the results depend decisively on the quality of the variational wavefunction — everything is completely pre-determined by the choice of trial wavefunction (and the physical arguments that motivate/define it).

In the particular case in which the trial wavefunction coincides with the exact ground state, the matrix elements for $A = H$ in (**) are all equal to E_0

$$E_x = \frac{\langle x | \mathcal{H} | \psi_\alpha \rangle}{\langle x | \psi_\alpha \rangle} = E_0$$

This property is called "zero variance", the more the wave function resembles the actual ground state, the more rapidly the variational energy converges with the number of iterations.

We could thus use this variance measure as an alternative criterion to optimize our variational approach.

Optimizing a trial wavefunction

So far we have seen how to employ Monte Carlo techniques to extract information (such as the energy) from a given trial wave function $|\psi_\alpha\rangle$.

Now we want to proceed to the next step and determine the set of parameters α that produces the "best" wavefunction within this class of wave functions parametrized by $\{\alpha\}$.

Local optimization of deterministic functions is something that has been worked on for a long time. VMC, though, is a stochastic process and returns the energy $E \pm \delta E$ with some error bar δE.

As a start, we would like to write down an approximation for the energy that is a deterministic function. Then, we could naively use a typical black box optimization method

To accomplish this, imagine we have $N = 1000$ configurations $\{x_1, x_2, \ldots, x_{N=1000}\}$ that are sampled from the distribution $|\psi_{\alpha_0}|^2$. We know that an approximation to the energy $E(\alpha_0)$ is

$$\langle E(\alpha_0) \rangle = \sum_{\{x_i\}} p_x^{(\alpha_0)} \cdot E_x^{(\alpha_0)}$$

It is also possible, though, that from the <u>same</u> $N = 1000$ configurations, we can also estimate the energy $E(\alpha_1)$ where α_1 is a different set of parameters then α_0.

Notice, that the configurations we want to use have <u>not</u> been sampled from $|\psi_{\alpha_1}|^2$ but $|\psi_{\alpha_0}|^2$. Nonetheless, we can <u>reweight</u> them so they look as if they have been sampled from this other distribution which we can accomplish via

$$\langle E(\alpha_1) \rangle \approx \frac{\sum_{\{x_i\}} \frac{P_x(\alpha_1)}{P_x(\alpha_0)} \cdot E_x^{(\alpha_1)}}{\sum_{\{x_i\}} \frac{P_x(\alpha_1)}{P_x(\alpha_0)}}$$

which gives us an estimate for the energy of our system with parameters $\{\alpha_1\}$.

Notice that $\langle E(\alpha) \rangle$ is now a <u>deterministic</u> function and we might be tempted to stick this function as an objective function into a black-box optimizer.

There are two problems with this approach though:

- If the wavefunctions $|\psi(\alpha_1)\rangle$ and $|\psi(\alpha_0)\rangle$ differ significantly, our average energy will become rather unreliable. We can overcome this by (occasionally) refreshing the set of configurations $\{x_i\}$.

- If we consider the limit of extremely small configuration sets, i.e. $N \to 1$, we will almost certainly find parameters $\{\alpha\}$ that are good for this (small) configuration set, but bad in general.

  This problem persists even in the case of moderately large configuration sets.

  The fundamental reason for this problem is that the energy is <u>not</u> bounded from below – consequently there are parameter regions that can (for a finite set of configurations) severly underestimate the energy. This is also called "undersampling".

The way out of this is a bag of tricks that include setting up an optimization criterion which is a combination of energy and its variance, regularly updating the configuration set $\{x_i\}$ and (relatively) sophisticated ways to determine gradients in the variational space.