Quantum Sampling

Dan Sehayek

October 12, 2018

Contents

1 Background

1.1 Density Matrices

Consider a quantum state $|\psi\rangle$. The **density matrix** ρ is defined such that $\rho = |\psi\rangle\langle\psi|$. The diagonal entries of ρ give the probabilities of the different possible states. Thus if there are N qubits then for a spin 1/2 system then there are 2^N diagonal entries and $\mathrm{Tr}(\rho) = 1$. For all of the following examples we will consider a spin 1/2 system.

Example: Suppose that $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. Then

$$\rho = |\psi\rangle\langle\psi| = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}\begin{pmatrix}\bar{\alpha} & \bar{\beta}\end{pmatrix} = \begin{pmatrix}\alpha\bar{\alpha} & \alpha\bar{\beta}\\ \alpha\bar{\beta} & \beta\bar{\beta}\end{pmatrix} = \begin{pmatrix}|\alpha|^2 & \alpha\bar{\beta}\\ \alpha\bar{\beta} & |\beta|^2\end{pmatrix} = \begin{pmatrix}P(0) & \cdot\\ \cdot & P(1)\end{pmatrix}$$

Example: Suppose that $|\psi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle$. Then

$$\rho = |\psi\rangle\langle\psi| = \begin{pmatrix}\alpha\\ \beta\\ \gamma\\ \delta\end{pmatrix}\begin{pmatrix}\alpha & \beta & \gamma & \delta\end{pmatrix} = \begin{pmatrix}|\alpha|^2 & \cdot & \cdot & \cdot\\ \cdot & |\beta|^2 & \cdot & \cdot\\ \cdot & \cdot & |\gamma|^2 & \cdot\\ \cdot & \cdot & \cdot & |\delta|^2\end{pmatrix} = \begin{pmatrix}P(00) & \cdot & \cdot & \cdot\\ \cdot & P(01) & \cdot & \cdot\\ \cdot & \cdot & P(10) & \cdot\\ \cdot & \cdot & \cdot & P(11)\end{pmatrix}$$

Note that the off diagonal entries are referred to as the **coherences** and contain information about the entanglement of the system.

1.2 Reduced Density Matrices

The **reduced density matrix** is defined as the **partial trace** of the density matrix. They are useful for measuring subsets of a given system. Consider for instance that we had the state $|\psi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle$ and that we solely wanted to measure the first qubit A while ignoring the second qubit B. Then our reduced density matrix would be

$$\rho^{(A)} = \mathrm{Tr}_B(\rho) = \begin{pmatrix}P(A=0) & \cdot\\ \cdot & P(A=1)\end{pmatrix}$$

If we have the tensor product $O_A \otimes O_B$ then the partial trace Tr_B is defined such that $\mathrm{Tr}_B(O_A \otimes O_B) = O_A\mathrm{Tr}(O_B)$. An example of this is given below.

Example: Consider $|\psi\rangle = \alpha|00\rangle + \beta|11\rangle = \alpha|0_A0_B\rangle + \beta|1_A1_B\rangle = \alpha|0_A\rangle|0_B\rangle + \beta|1_A\rangle|1_B\rangle$. We would like to determine $\rho^{(A)}$ so that we can obtain the probability of measuring qubit A to be 0. Our first step would be to compute ρ.

$$\begin{aligned}\rho = |\psi\rangle\langle\psi| &= |\alpha|^2|00\rangle\langle 00| + \alpha\beta^*|00\rangle\langle 11| + \alpha^*\beta|11\rangle\langle 00| + |\beta|^2|11\rangle\langle 11|\\ &= |\alpha|^2|0_A\rangle\langle 0_A| \otimes |0_B\rangle\langle 0_B| + \alpha\beta^*|0_A\rangle\langle 1_A| \otimes |0_B\rangle\langle 1_B|\\ &+ \alpha^*\beta|1_A\rangle\langle 0_A| \otimes |1_B\rangle\langle 0_B| + |\beta|^2|1_A\rangle\langle 1_A| \otimes |1_B\rangle\langle 1_B|\end{aligned}$$

The partial trace can then be computed as follows.

$$\begin{aligned}
\mathrm{Tr}_B(\rho) &= |\alpha|^2 \,\mathrm{Tr}_B(|0_A\rangle\langle 0_A| \otimes |0_B\rangle\langle 0_B|) + \alpha\beta^* \mathrm{Tr}_B(|0_A\rangle\langle 1_A| \otimes |0_B\rangle\langle 1_B|) \\
&+ \alpha^*\beta \mathrm{Tr}_B(|1_A\rangle\langle 0_A| \otimes |1_B\rangle\langle 0_B|) + |\beta|^2 \,\mathrm{Tr}_B(|1_A\rangle\langle 1_A| \otimes |1_B\rangle\langle 1_B|) \\
&= |\alpha|^2 \,|0_A\rangle\langle 0_A| \,\mathrm{Tr}_B(|0_B\rangle\langle 0_B|) + \alpha\beta^* \,|0_A\rangle\langle 1_A| \,\mathrm{Tr}_B(|0_B\rangle\langle 1_B|) \\
&+ \alpha^*\beta \,|1_A\rangle\langle 0_A| \,\mathrm{Tr}_B(|1_B\rangle\langle 0_B|) + |\beta|^2 \,|1_A\rangle\langle 1_A| \,\mathrm{Tr}_B(|1_B\rangle\langle 1_B|) \\
&= |\alpha|^2 \,|0_A\rangle\langle 0_A| + |\beta|^2 \,|1_A\rangle\langle 1_A|
\end{aligned}$$

Expanding gives:

$$\rho^{(A)} = \mathrm{Tr}_B(\rho) = \begin{pmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{pmatrix} = \begin{pmatrix} P(A=0) & 0 \\ 0 & P(A=1) \end{pmatrix}$$

Thus we can see that the reduced density matrix $\rho^{(A)} = \mathrm{Tr}_B$ gives us the probabilities of measuring 0 and 1 for qubit A while ignoring qubit B.

2 Perfect Sampling

2.1 The Algorithm

Recall the previous example. Let us say that we used $P(A = 0)$ from our reduced density matrix to sample qubit A. We can denote the measured state $\tilde{A}$. If we then wanted to sample qubit B, we would want to calculate the reduced density matrix $\rho^{(B)}$ given $\tilde{A}$. We can denote this $\rho^{(B)}(\tilde{A})$. This reduced density matrix can be computed by first projecting our measurement onto ρ and then normalizing based on our previous measurement.

$$\rho^{(B)}(\tilde{A}) = \frac{\left\langle \tilde{A}|\psi \right\rangle \left\langle \psi|\tilde{A} \right\rangle}{P(\tilde{A})}$$

The different possibilities for the numerator are:

$$\langle 0| \left(|\alpha|^2 \,|00\rangle\langle 00| + \alpha\beta^* \,|00\rangle\langle 11| + \alpha^*\beta \,|11\rangle\langle 00| + |\beta|^2 \,|11\rangle\langle 11|\right) |0\rangle = |\alpha|^2 \,|0\rangle\langle 0|$$

$$\langle 1| \left(|\alpha|^2 \,|00\rangle\langle 00| + \alpha\beta^* \,|00\rangle\langle 11| + \alpha^*\beta \,|11\rangle\langle 00| + |\beta|^2 \,|11\rangle\langle 11|\right) |1\rangle = |\beta|^2 \,|1\rangle\langle 1|$$

Thus the different possible matrices for the reduced density matrix are

$$\rho^{(B)}(0) = \frac{1}{|\alpha|^2} \begin{pmatrix} |\alpha|^2 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\rho^{(B)}(1) = \frac{1}{|\beta|^2} \begin{pmatrix} 0 & 0 \\ 0 & |\beta|^2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

If we measured qubit A to be 0 then there should be a 100% probability that we will measure qubit B to be 0 and likewise for 1. Thus this result makes sense.

Let us now consider a slightly more complicated system of 4 qubits. Let us suppose that we would like to first sample the fourth qubit. Then we would need the following reduced density matrix.

$$\rho^{(4)} = \mathrm{Tr}_{\sigma_1\sigma_2\sigma_3}(|\psi\rangle\langle\psi|)$$

Once we obtained this density matrix, we would have the probability of the fourth qubit being 0. We can use this probability to sample/measure σ_4 using a random number generator. Let us call the measured state $\tilde{\sigma}_4$. Our next step would be to sample the third qubit. Thus we would need $\rho^{(3)}$ given $\tilde{\sigma}_4$.

$$\rho^{(3)}(\tilde{\sigma}_4) = \frac{\mathrm{Tr}_{\sigma_1\sigma_2}\left(\langle\tilde{\sigma}_4|\psi\rangle\langle\psi|\tilde{\sigma}_4\rangle\right)}{P(\tilde{\sigma}_4)}$$

Notice once again that we have projected our measured state onto $|\psi\rangle\langle\psi|$ and that we have normalized our reduced density matrix based on the result of the previous measurement. Since we have four qubits this time, we still need to take the partial trace with respect to the first two qubits. Once again we can sample our current qubit based on the probability from $\rho^{(3)}$. Let us call the measured state $\tilde{\sigma}_3$. Next we need $\rho^{(2)}$ given $\tilde{\sigma}_3$ and $\tilde{\sigma}_4$.

$$\rho^{(2)}(\tilde{\sigma}_3, \tilde{\sigma}_4) = \frac{\mathrm{Tr}_{\sigma_1}(\langle\tilde{\sigma}_3\tilde{\sigma}_4|\psi\rangle\langle\psi|\tilde{\sigma}_3\tilde{\sigma}_4\rangle)}{P(\tilde{\sigma}_3, \tilde{\sigma}_4)}$$

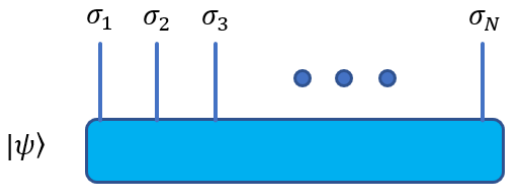
Notice here that we need the probability of both of the previous measurements. This is given by $P(\tilde{\sigma}_4)\cdot P(\tilde{\sigma}_3|\tilde{\sigma}_4)$ where the second term represents a conditional probability. Once again we can use this to sample the second qubit and then obtain $\rho^{(1)}$.

$$\rho^{(1)}(\tilde{\sigma}_2, \tilde{\sigma}_3, \tilde{\sigma}_4) = \frac{\langle\tilde{\sigma}_2\tilde{\sigma}_3\tilde{\sigma}_4|\psi\rangle\langle\psi|\tilde{\sigma}_2\tilde{\sigma}_3\tilde{\sigma}_4\rangle}{P(\tilde{\sigma}_2, \tilde{\sigma}_3, \tilde{\sigma}_4)}$$

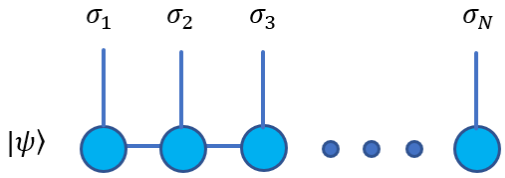
This time our normalization term would be given by $P(\tilde{\sigma}_4)\cdot P(\tilde{\sigma}_3|\tilde{\sigma}_4)\cdot P(\tilde{\sigma}_2|\tilde{\sigma}_3\tilde{\sigma}_4)$. Finally we can use this to measure our final qubit and obtain a full sample!

2.2 Implementation using Tensor Networks

Recall the tensor network representation of a quantum state ψ of N qubits.

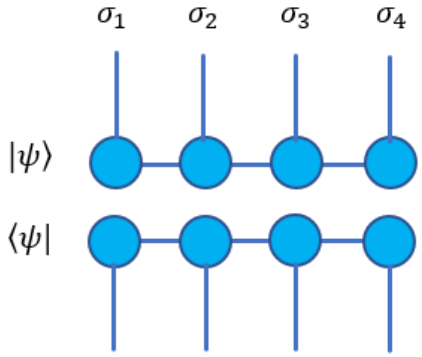


Each line represents a **site index**. In general a quantum system with N qubits will be a rank N tensor. Specifying a value for each site index will give us the corresponding amplitude coefficient. In the case of a spin 1/2 system, the possible values for each site index will be 0 and 1. If we wanted to represent $|\psi\rangle$ in **Matrix Product State** or **MPS** form we would use the following schematic.

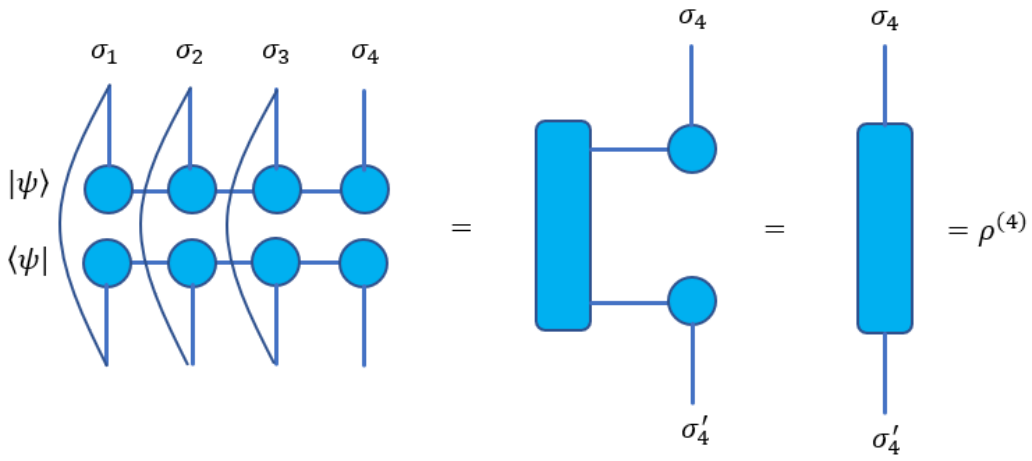


Here $|\psi\rangle$ is now represented as a product of matrices. Each circle represents a matrix and has one or two **bond indices** in addition to its site index. If one wanted to obtain the original form, they would simply contract over all of these indices which correspond to multiplying all of the matrices.

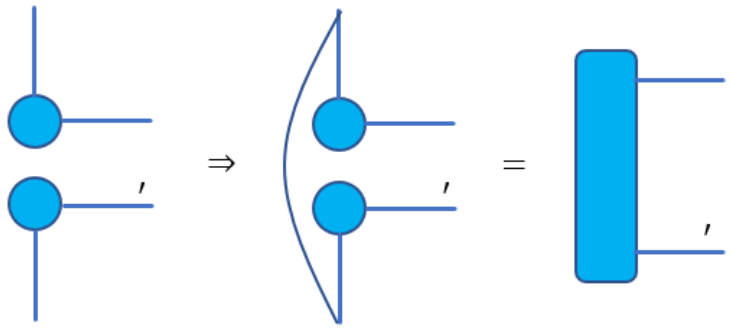
For the perfect sampling algorithm, we first need to represent the density matrix ρ. Let us consider the case of $N = 4$. Our density matrix can be represented as follows.



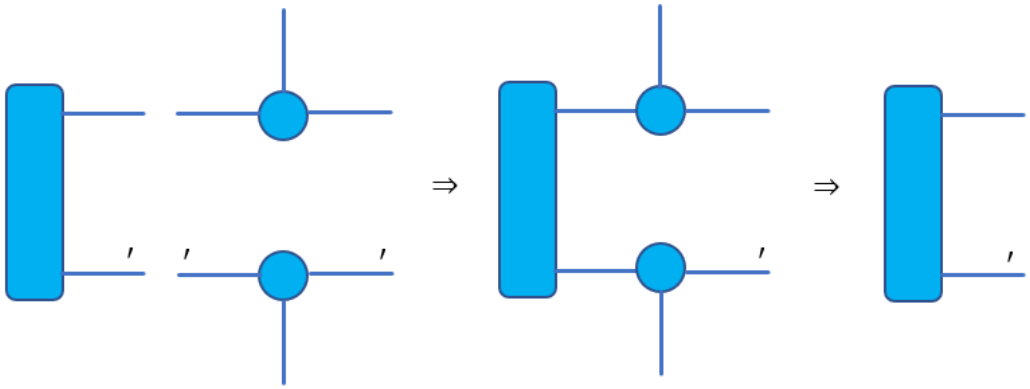
Recall the for our first measurement we require $\rho^{(4)} = \mathrm{Tr}_{\sigma_1\sigma_2\sigma_3}(\rho)$. This partial trace corresponds to contracting the first three matrices as follows.



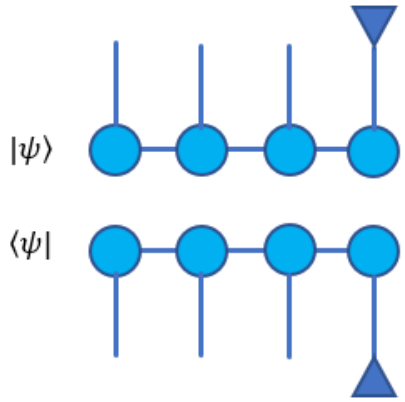
The first set of matrices is contracted about the site index. The second set of matrices is contracted with the first by priming the bond indices of the bra. In general, **priming** an index of a given tensor will ensure that contraction does not occur over that index when multiplied by another tensor with an equivalent index. The diagram below shows the contraction for the first set of tensors.



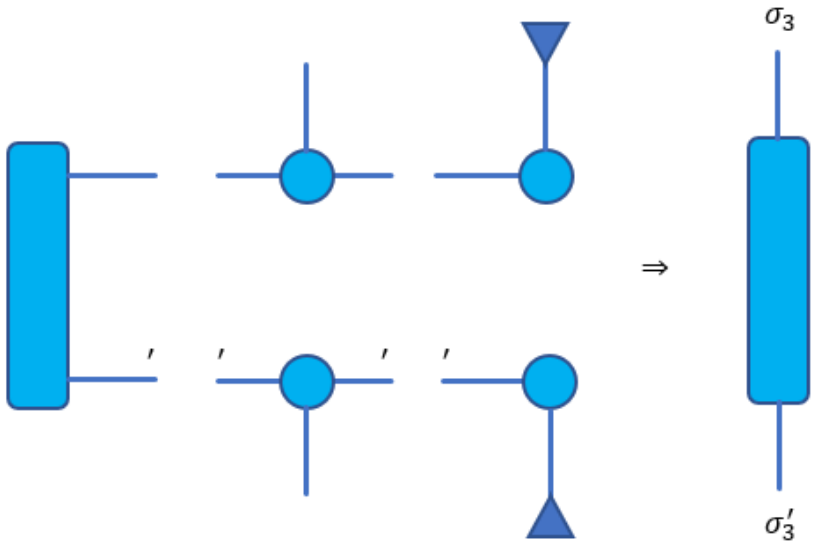
Before contracting we prime the bond index of the ket to ensure that contraction does not occur over this index. This overall step represents the partial trace over σ_1. In ITensor this is done as `psi.A(1) * dag(prime(psi.A(1),bondIndex1))`. Our next step is to contract over the second set of tensors. This is shown in the following diagram.



This represents the partial trace over σ_1 and σ_2. Note that in the last step we contracted over the site index of σ_2. A similar process is done for the third set of tensors. Essentially at each step we prime both bond indices of the ket and then contract the bond indices followed by the site indices. Once we have completed these steps, we are left with $\rho^{(4)} = \mathrm{Tr}_{\sigma_1\sigma_2\sigma_3}(\rho)$. We can extract the first entry of $\rho^{(4)}$ to obtain the probability that the fourth qubit will become the 0 state. We can then use this to obtain our sample $\tilde{\sigma}_4$ which can be obtained using a random number generator. For our next measurement, we need $\rho^{(3)}(\tilde{\sigma}_4) = \mathrm{Tr}_{\sigma_1\sigma_2}(\langle\tilde{\sigma}_4|\psi\rangle\langle\psi|\tilde{\sigma}_4\rangle)$ divided by our normalization factor $P(\tilde{\sigma}_4)$. To denote the projection of the state $|\tilde{\sigma}_4\rangle$ onto our density matrix, we use the following notation.



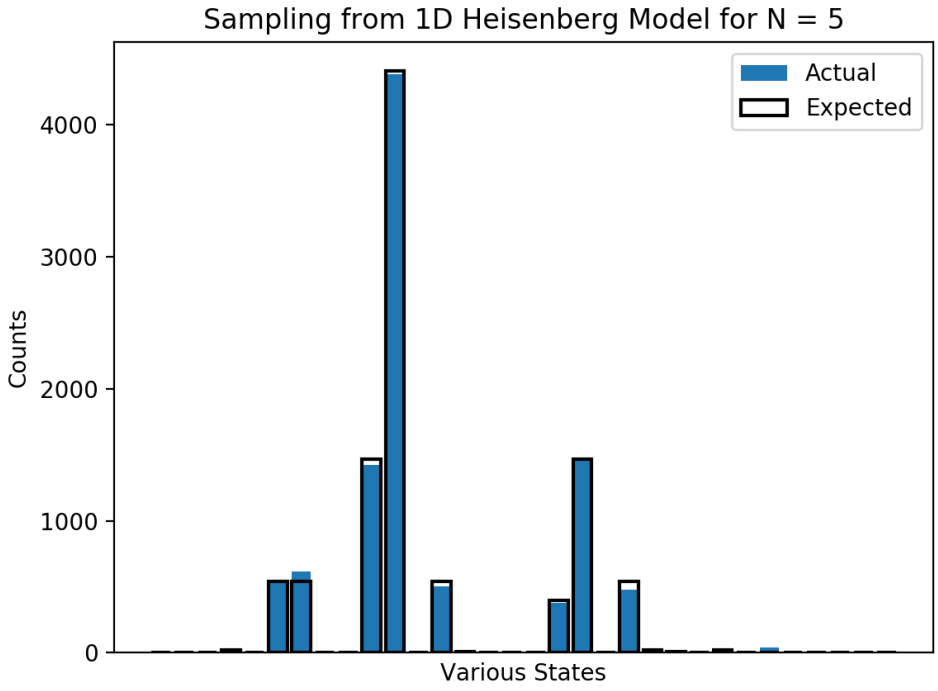
At this point, we contract the first two sets of tensors as we previously did and then contract with the third set while priming the site index on the bra. Finally we contract with the fourth set of tensors onto which the measurement state $\tilde{\sigma}_4$ was projected.



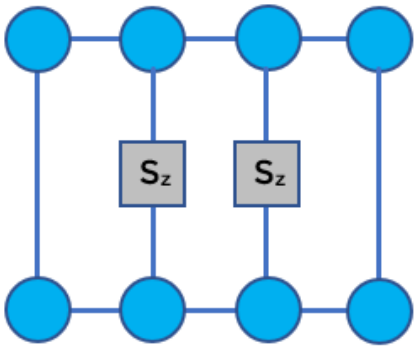
Once again we can extract $P(\sigma_3 = 0|\tilde{\sigma}_4)$ and use this to obtain the next measurement $\tilde{\sigma}_3$. A similar can be repeated for our next two measurements to obtain $\tilde{\sigma}_2$ and $\tilde{\sigma}_1$. We will then have obtained a complete sample! When writing the script for this algorithm, it is advised to store the contracted states as it would inefficient to recalculate them at every iteration.

2.3 Validation Techniques

Once the perfect sampling routine is executed for N qubits, it would obviously be nice to have a method of verifying that our simulator is indeed working correctly. In the case where N is small enough such that it is feasible to store all of the amplitude coefficients, one can plot a histogram to compare the observed frequencies with the expected frequencies given by $M \cdot |\psi(\sigma)|^2$ where M is the number of samples. An example is shown below for the 1D Heisenberg model with $N = 5$.



What about for large N? In this case we can compare the expected value of an observable with the value from the samples. Our two main possibilities are **diagonal** observables and **off diagonal** observables. In terms of calculating the expected value, this can be done using tensor networks and is done similarly for both cases. Let us consider the two spin observable $S_2^z S_3^z$. The expected value of such an observable given by $\langle\psi| S_2^z S_3^z |\psi\rangle$ can be calculated using the following tensor network.



To construct this tensor network using ITensor, the following code was used. Similar to our routine for perfect sampling, priming was essentially used to control the contractions while dag was used to conjugate certain tensors.

```
// Here we measure a two spin operator correlation function and compare
// with the value obtained from sampling. We will use the second and
// third spin operators.
auto Sz2 = sites.op("Sz",2);
auto Sz3 = sites.op("Sz",3);

// Place the orthogonality center on site 2
// This completes the contractions for us.
psi.position(2);

// Contract second ket tensor with operator
ITensor C = psi.A(2);
C *= Sz2;

// Contract with second bra tensor while making sure to prime the site index
// as well as the bond index that is common with the following tensor
auto CI1 = commonIndex(psi.A(2),psi.A(3),Link);
C *= dag(prime(prime(psi.A(2),Site),CI1));

// Repeat similar process for third set of tensors
C *= psi.A(3);
C *= Sz3;
auto CI2 = commonIndex(psi.A(3),psi.A(2),Link);
C *= dag(prime(prime(psi.A(3),Site),CI2));

// Result corresponding to <psi|S2S3|psi>
auto result = C.real();
```

A similar approach can be taken for an off diagonal observable. For the Hamiltonian H, one can simply write `overlap(psi,H,psi)` where `psi` is the MPS state and thus can obtain the expected value is one line. For the 1D Heisenberg model, the code specifying the Hamiltonian and the MPS state is as follows.

```
// Create 1d Heisenberg Hamiltonian
auto ampo = AutoMPO(sites);
for(int j = 1; j < N; ++j)
    {
    ampo += 0.5,"S+",j,"S-",j+1;
    ampo += 0.5,"S-",j,"S+",j+1;
    ampo +=   1,"Sz",j,"Sz",j+1;
    }
auto H = MPO(ampo);

// Set up random initial wavefunction
auto psi = MPS(sites);
```

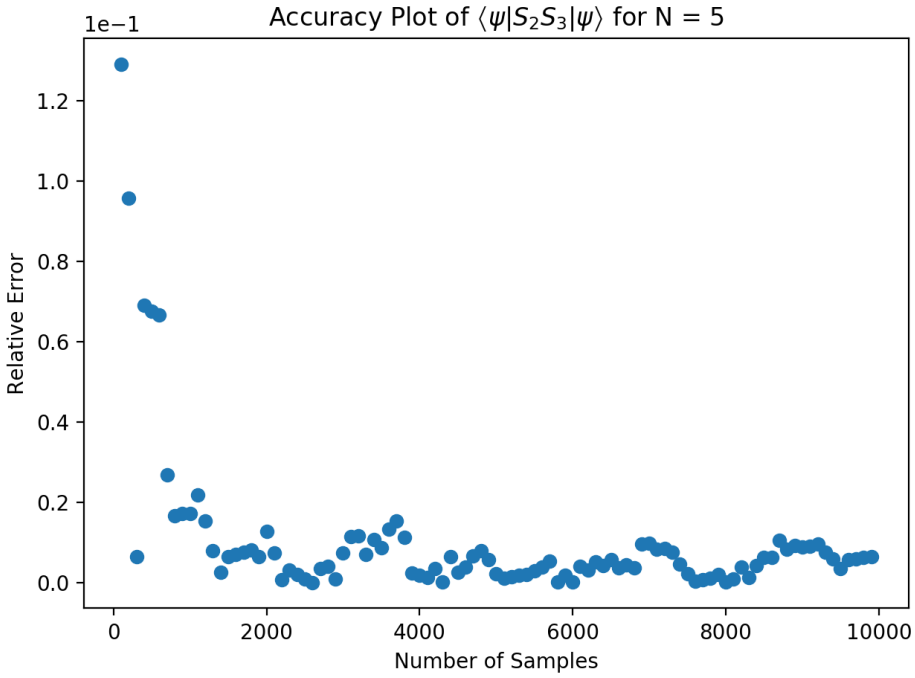
The Hamiltonian corresponding to the code above is

$$H = \sum_{j=1}^{N} S_j^+ S_{j+1}^- + S_j^- S_{j+1}^+ + S_j^z S_{j+1}^z$$

MPO is the class for storing matrix product operators and **MPS** is the class for storing matrix product states. Our next step is to calculate the average value of the observable from the samples. The value of a diagonal observable O corresponding to a sample σ_k can be calculated simply by determining the corresponding diagonal entry. For instance, if our sample state is 0101 and our observable is $S_2^z S_3^z$ then the value of the observable corresponding to this state will be $\langle 1|S^z|1\rangle\langle 0|S^z|0\rangle = (-0.5)(0.5) = -0.25$. Overall our expression for the measured value of the observable from the samples will be as follows.

$$\langle O\rangle \approx \frac{1}{M}\sum_{k=1}^{M} O_{\sigma_k}$$

One can plot a graph of the relative error versus the number of samples to help verify their simulation. This is particularly useful in determining a minimum number of samples to use before attempting to train an RBM. A recommended threshold value for the relative error would be 0.001 but this will vary. An example of this graph is shown below for $N = 5$.



For an off diagonal observable, one must use the **local estimator** given by $O_{\sigma_k}^{[L]}$ as we now must take into account the off diagonal entries. The value of the local estimator for a particular spin configuration is given below.

$$O_{\sigma_k}^{[L]} = \sum_{\sigma'} \frac{\psi(\sigma')}{\psi(\sigma)} O_{\sigma\sigma'}$$

We can then once again obtain our measured value for the observable as follows.

$$\langle O \rangle \approx \frac{1}{M} \sum_{k=1}^{M} O_{\sigma_k}^{[L]}$$

As an example, let us consider the $S_j^+ S_{j+1}^-$ operator. This corresponds to applying the raising operation to the j^{th} qubit and the lowering operation to the following qubit. Recall that the raising operator corresponds to $S^+|0\rangle = |1\rangle$ and $S^+|1\rangle = 0$ and the lowering operation corresponds to $S^-|0\rangle = 0$ and $S^-|1\rangle = |0\rangle$. Our local estimator value will then be equivalent to the amplitude of the transformed state divided by the amplitude of the original state. This is demonstrated below.

$00... \Rightarrow 0$

$01... \Rightarrow \psi(10...)/\psi(01...)$

$10... \Rightarrow 0$

$11... \Rightarrow 0$

For our second term we have $O_{\sigma\sigma'} = 1$ while for all other terms we have $O_{\sigma\sigma'} = 0$. Once again we can plot a graph of the relative error versus the number of samples. The graph shown below is for $N = 5$.