Simulated Annealing & Iterative Improvement on the Ising Model

Team 4 **Natali Alfonso, Simon Kern, Vangelis Kostas**

Machine Learning (NWI-NM048C)

Prof. B. Kappen Due date: 31/01/2017

Introduction

Please run "Run_me.m" in the "Ising" folder

We have the following combinatoric optimization problem:

$$E = -\frac{1}{2}x^t w x$$

We notice a similarity with the energy function of an Ising Model where the applied field $H = 0$. An Ising model is an array of spins (e.g., atoms that can take states 1) that are magnetically coupled to each other. If one spin is, say, in the +1 state then it is energetically favourable for its immediate neighbours to be in the same state, in the case of a ferromagnetic model, and in the opposite state, in the case of an antiferromagnet.
If x is a binary vector so $x_i = +/- 1$ then finding a minimum for E is intractable. **For real values of x and $\|x\| = 1$ we could minimize the value of the energy with $\nabla E = 0$ with respect to x.**

For w_{ij} that are all positive this global minimization problem turns into a minimization of all the spin terms separately, this would be a **ferromagnetic** system. **Since in that case all the terms would add up towards one direction $E = \frac{1}{2}\sum_{ij} w_{ij}x_i x_j$ so minimizing $w_{ij}x_i x_j$ which in this case would mean $x_i = x_j$ for all i, j.**

In the case where $w_i j$ take random binary values though not all terms can be minimized simultaneously and instead there needs to be a global compromise, this would be a **frustrated** system.

1 Iterative Improvement

We decided to make an optimization on the code so we dont have to calculate the energy function every time but just the energy change.

$$
\begin{aligned}
u &= x - y \\
u'wu &= x'wx - 2x'wx_{new} + x'_{new}wx_{new} \\
u'wu &= x'wx - 2x'wx_{new} + x'_{new}wx_{new} \\
u'wu - x'wx + 2x'wx_{new} &= x'_{new}wx_{new} \\
u'wu + x'wx - 2x'wx + 2x'wx_{new} &= x'_{new}wx_{new} \\
u'wu + x'wx + 2x'w(x_{new} - 2x) &= x'_{new}wx_{new} \\
x'w(x_{new} - 2x) = x'w(u - x) &= x'wu - x'wx \\
u'wu + x'wx + 2x'w(x_{new} - 2x) &= x'_{new}wx_{new} \\
u'wu + x'wx + 2x'wu - 2x'w'x &= x'_{new}wx_{new} \\
u'wu - x'wx + 2x'wu &= x'_{new}wx_{new}
\end{aligned}
$$

Where $x'wx$ is known and the remaining $u'wu$ and $2x'wu$ have a factor of n less calculations because they are just the multiplication of a column from w by scalars. **So for an nxn weight matrix w we will be making a factor of n less computations.**
This can be clearly seen in our decrease in execution time for each loop below.

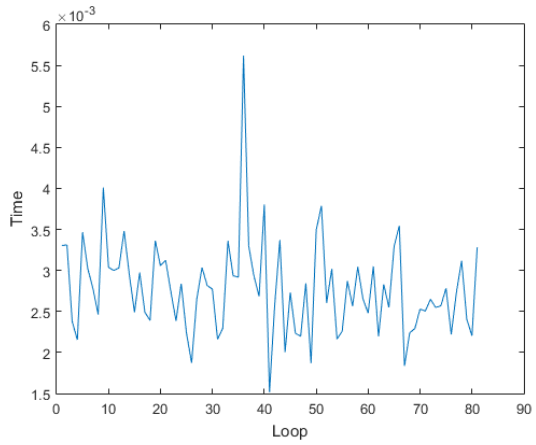

Figure 1: Optimized code 100 spin system.

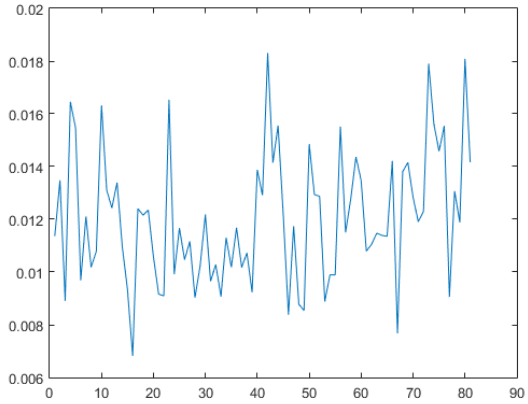

Figure 2: Un-optimized code 100 spin system.

As we can see for a 100 spin ferromagnetic system (10x10 quadratic grid) we get a speed up from 130 milliseconds to 2.5 milliseconds, almost **10 times as fast**(less than that because matlab already optimizes the matrix multiplication). The result we expected from our theoretical calculations.

1.1 Compare the ferro-magnetic and frustrated problems. How many restarts are needed for reproducible results?

We will call the results reproducible if we get the same minimum energy for 2 separate random initiations of the iterative algorithm. If we reproduce the minimum energy we stop the iterations. The following tests will be done on **quadratic** grids of spins.

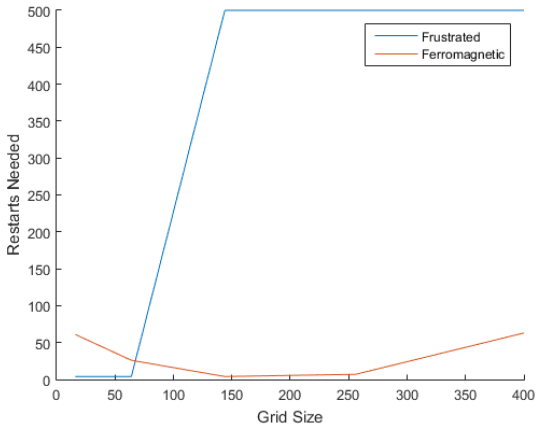

Figure 3: Number of restarts.

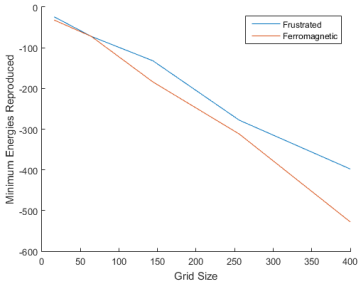

Figure 4: Minimum energies found per size.

As we can see in Figure 3 the number of restarts for the frustrated system far outweighs that for the ferromagnetic one. In fact we see that it reaches our maximum of 500 which means that we can not find a consistent minimum energy even after 500 restarts and we stop the iterations prematurely.
This result is expected since a simple iterative change of the spins will find it much easier to find a solution when subsequent changes do not ruin the advancement made on the previous changes. That is clearly not true for the frustrated system.

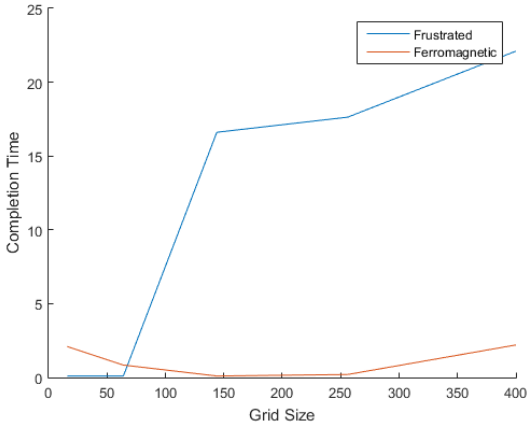

Figure 5: Completion times until predictability.

In Figure 4 we see the minimum energies found follow a almost linear relation with the size of the grid. Figure 5 shows the completion times which skyrocket for the frustrated system after a size of about 100 even though at those sizes the iteration is prematurely stopped, so the iteration would go on even longer if we allowed it.

1.2 For the frustrated problem, study the influence of the neighborhood size on the quality of the solution and the cpu time required.

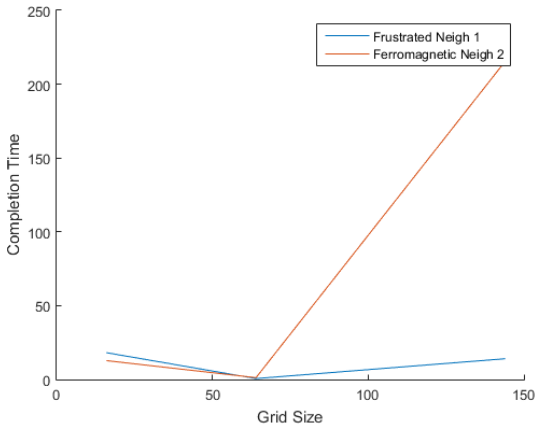

Figure 6: Number of restarts.

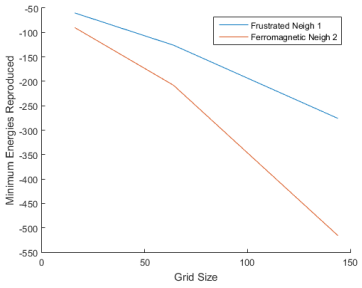

Figure 7: Minimum energies found per size.

As we can see in Figure 7 the quality of the solutions for a neighborhood of 2 spins is consistently better for every size of grid tested. On the other hand in Figure 6 the completion time rises exponentially with the size of the grid and becomes factors of magnitude larger than the algorithm with a neighborhood of 1 spin. These results are expected when taking into account the size difference that one extra spin introduces into the possible combinations for each change. We didnt test larger grids because the algorithm would take too long to complete.

2 Simulated Annealing

Simulated annealing is an advanced method for combinatoric optimization. We instead sample from a probability distribution with an adjustable parameter β that gets bigger on after every sampling process and is in effect making the maximums of the $\epsilon^{-\beta E(x)}$ function more and more pronounced. So at the end of every stage of simulated annealing we get a good estimation of the mean state to start the next stage with.
The probability distribution we will be sampling from is:

$$p(x) = \frac{\exp^{-\beta E(x)}}{Z}$$

With Z the intractable normalizing factor. For our sampling step we will be using Metropolis-Hastings.
Which is a Markov chain Monte Carlo (MCMC) method for obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult.
For each Markov Chain sampled we will be estimating the variance and mean, and we will only stop when the variance of the samples becomes zero which means we have always sampled the same state, the minimum energy state.

$$\langle E \rangle = \frac{1}{T1} \sum_{i=1}^{T1} E(x_i)$$

$$\langle \sigma^2 \rangle = \frac{1}{T1} \sum_{i=1}^{T1} (E(x_i) - \langle E \rangle)^2$$

Where $T1$ is the length of the Markov chain.

2.1 Reproduce parts of figure 31.11 of MacKay a ferromagnetic system of n = 50 spins, ie. estimate the mean energy and the standard deviation of the energy. Repeat this for a frustrated system by choosing random couplings.

To do this we can clearly not use the matrix w randomly initiated. Instead we will have to apply a neighborhood function on the w matrix in order to make it represent a triangular grid with continuous border conditions. So only the neighbouring spins will get a weight different than 0 in the weight matrix.

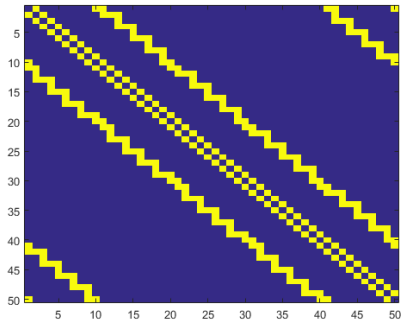

Figure 8: 50 spin **ferromagnetic** triangular grid weight matrix w

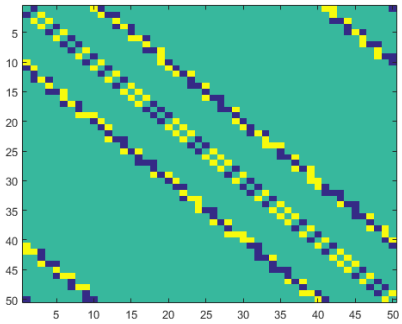

Figure 9: 50 spin **frustrated** triangular grid weight matrix w

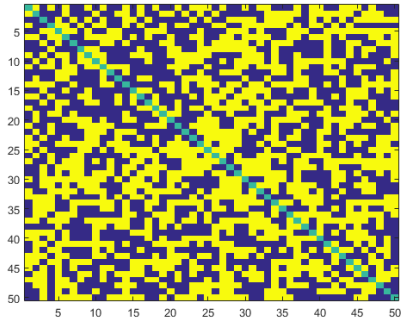

Figure 10: 50 spin random initiation of weight matrix w

As we can see after the application of our neighborhood function on the weight matrix it takes values different than zero only for spins that are in neighboring locations on the grid. We

can also create 1D grids and quadratic grids with that same function.

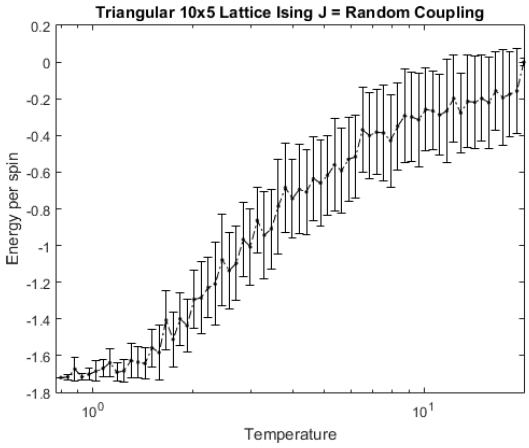

Figure 11: Energy graph for randomly coupled triangular grid.

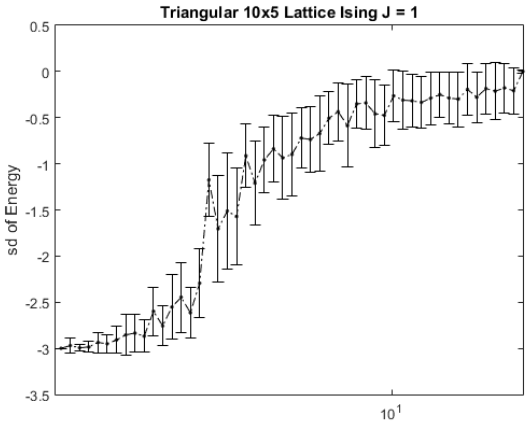

Figure 12: Energy graph for ferromagnetic triangular grid.

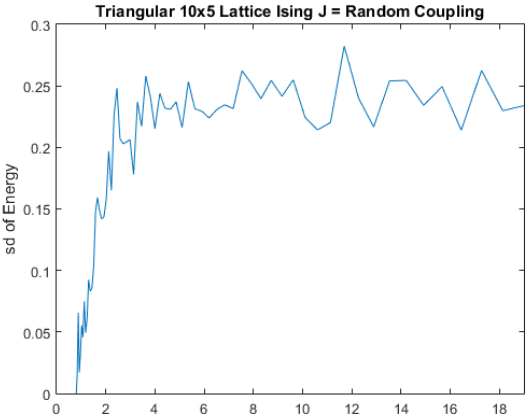

Figure 13: SD of Energy graph for randomly coupled triangular grid.

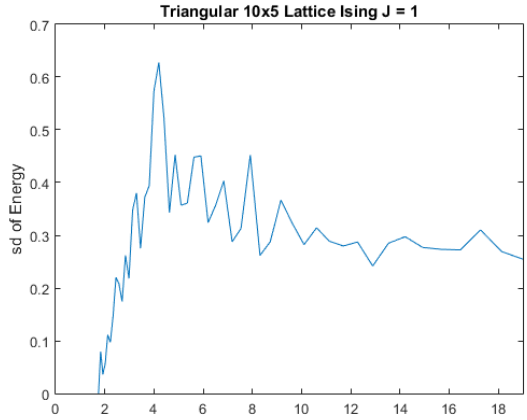

Figure 14: SD of Energy graph for ferromagnetic triangular grid.

As we can see in figure 13 the standard deviation of the energy doesnt really show a peak but it does show a fast decrease close to T=0. That means that this system doesnt show a very pronounced phase transition and we can see that in the Energy graph in figure 11. The Energy reaches the minimum on a gradual slope, less pronounced than the Ferromagnetic system but more so than the anti-ferromagnetic one.

The Ferromagnetic system in figures 12 and 14 shows the same characteristics as the one in figure 31.11 from the book. It reaches the minimum of **-3** per spin energy and it shows a peak in the standard deviation of energy when the phase transition occurs. In our case the error bars are way bigger than in the book since the book has a system with 4096 spins but we only have 50. We have a very good approximation of the critical temperature from Figure 14. We predict a critical temperature around **4.05** and the theoretical value for a triangular grid would be **3.6410** as given by Fisher (1967, p.671).

2.2 Study the effect of initial beta and the cooling schedule (factor) and the length of the Markov Chain T1 on the performance and reproducibility of the SA result. Estimate the critical temperature in both cases. Use a larger n to get more accurate results if your computer or patience allows.

For these simulations we will be using a ferromagnetic system of 300 spins with triangular grid.

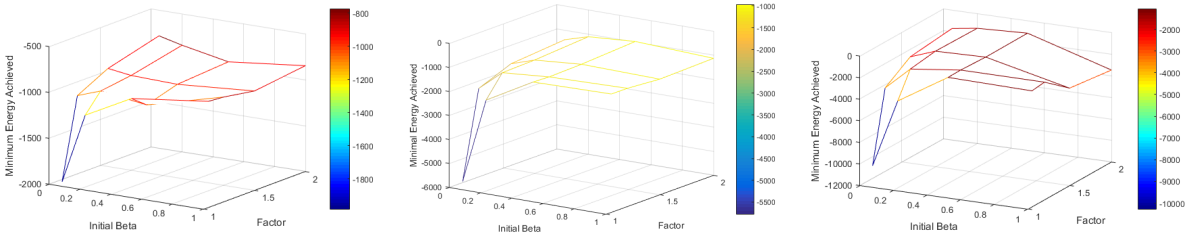

Figure 15: Minimum energy graph for different initial β and multiplication factors. For 200,1000 and 2000 Markov chain lengths.

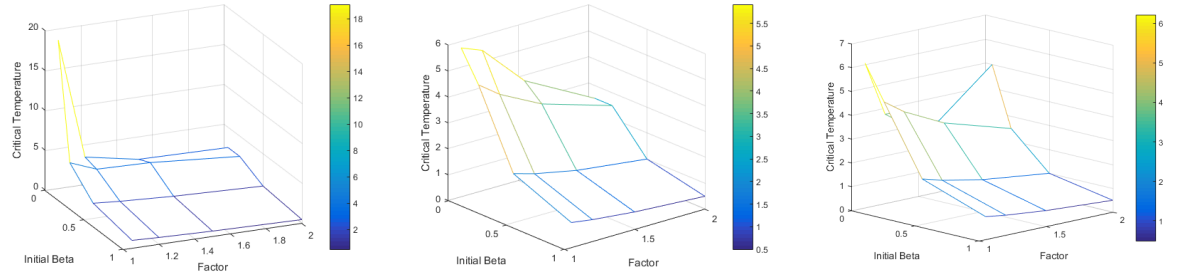

Figure 16: Critical Temperature graph for different initial β and multiplication factors. For 200,1000 and 2000 Markov chain lengths.

As we can see in figure 15 the initial β and the cooling schedule both play a role in the minimum E found with SA. We also note the difference from the first image of Figure 15 to the third one. With a longer Markov chain we get lower Energy minimums so better results as expected.
One interesting part of the images is the plateau they display. That shows that on that plateau we have high reproducibility but our performance isnt the best. It seems that we would have to choose between reproducibility and performance. A good balance would seem to arise for a $\beta \approx 0.2$ and cooling factor **1.1**
When we take a look at the estimated critical temperatures in Figure 16, we again see the importance of a long Markov chain. Taking into account that the theoretical value for the triangular grid is arround **3.64** we again see the closest approximations for the longer chains (16b,16c) and $\beta \approx 0.2$ and cooling schedule **1.2**

2.3 Which method (SA or Iter) has the best performance in terms of speed and quality?

For ferromagnetic systems the iterative improvement method is better in terms of quality and speed.That is expected since its good at maximizing all the individual terms and in the ferromagnetic case and thats all thats needed.

In any case other than the ferromagnetic one both the performance and speed is better for SA. Since SA doesnt change speed depending on the type of system(ferromagnet or not) but the Iter method will struggle to reproduce its minimum for big systems and non ferromagnetic ones.

2.4 Put n = 200 and make an instance with the randomseed fixed (rand(state,0)). Try to find the best solution and compare with your fellow students.

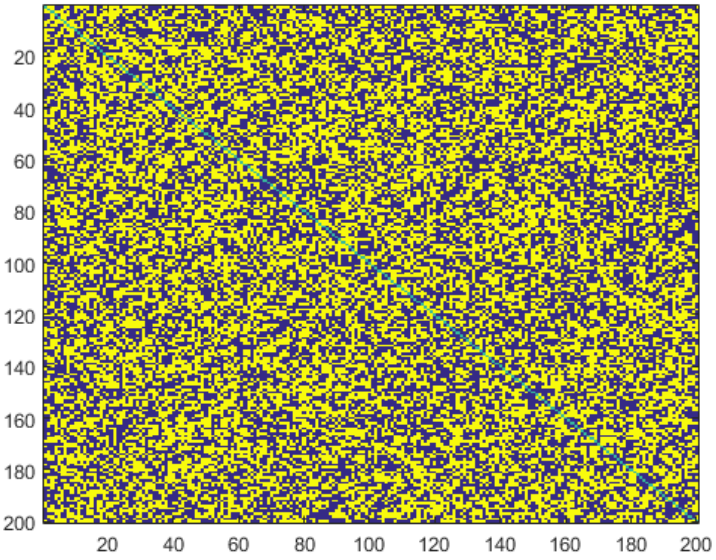

Figure 17: The initiated weight matrix as requested with the random seed set to 0.

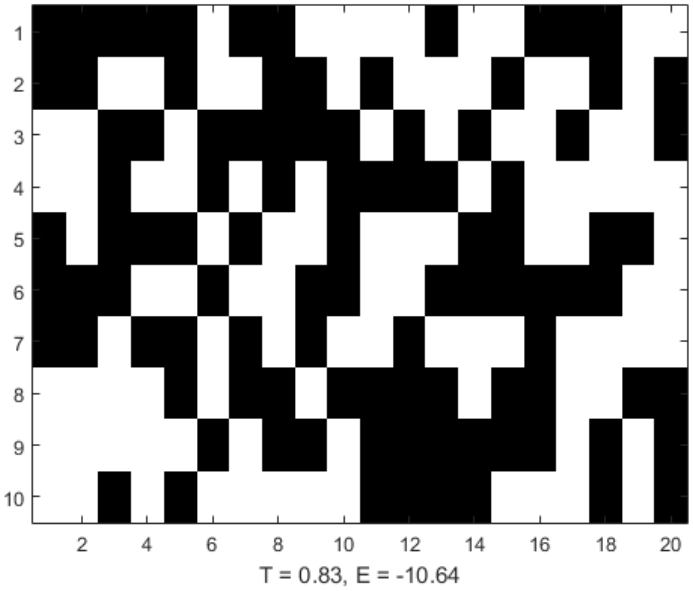

Figure 18: Final state of minimum energy.

We reached a minimum energy of -2128 and the final state can be seen in figure 18. That was reached in temperature 0.83 and per spin energy was -10.64