Real Space Renormalization and Applications

A Project Report

submitted by

KARTHIK V (EP13B012)

in partial fulfilment of the requirements
for the award of the degree of

BACHELOR OF TECHNOLOGY

DEPARTMENT OF PHYSICS
INDIAN INSTITUTE OF TECHNOLOGY MADRAS.

April 2017

THESIS CERTIFICATE

This is to certify that the thesis titled **Real Space Renormalization and Applications**, submitted by **Karthik V**, to the Indian Institute of Technology, Madras, for the award of the degree of **Bachelor of Technology**, is a bona fide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr.Ashwin Joy
Research Guide
Professor
Dept. of Physics
IIT-Madras, 600 036

Place: Chennai

Date: 27th April 2017

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to Dr.Ashwin Joy for guiding me throughout the project and giving me a first hand experience in research. I would also like to thank Dr.Dillip Satapathy for kindling my interest in Statistical Mechanics.

In addition, a thank you to Dr.Koushik Balasubramanian for motivating me to work on this project.

Finally, I would like to thank my parents and friends whose continued support helped me finish the project.

ABSTRACT

The concept of renormalization group(RG) theory has been extensively used to solve problems in condensed matter physics , phase transitions and high energy physics. RG is an iterative coarse-graining scheme that allows for the extraction of relevant features as a physical system is examined at different length scales. Here, we explore the possibility of applying RG concepts in deep learning. In recent times, a lot of deep learning techniques have yielded record-breaking results on a diverse set of difficult machine learning tasks in computer vision, speech recognition, and natural language processing. Despite this enormous success at feature extraction, there is very less theoretical understanding as to the reason behind this success. We attempt to shed light on some of the reasons behind the success of Restricted Boltzmann Machines(RBMs) using RG.

Chapter 1 gives a brief introduction to phase transitions in general and Ising model in particular. Chapter 2 introduces Renormalization Group Theory and deals with various attempts to solve the 2D Ising model problem using RG. We talk about the applications of RG in explaining the working of RBMs in chapter 3. We conclude the report by stating further work that could be done in this field.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

Phase Transitions in the Ising Model

A phase is a state of matter in thermodynamic equilibrium. A system could be in several different states or phases depending upon the macroscopic condition of the system. For example, ice , water and water vapour are three different phases of a collection of H_2O molecules. Depending on conditions such as pressure and temperature, the set of molecules exists in the phase with the lowest free energy. One way to find the presence of phase transitions is using Ehrenfest criteria.

Ehrenfest Criteria

For nth order phase transition, there is a finite discontinuity or divergence in the nth derivative of the free energy function.

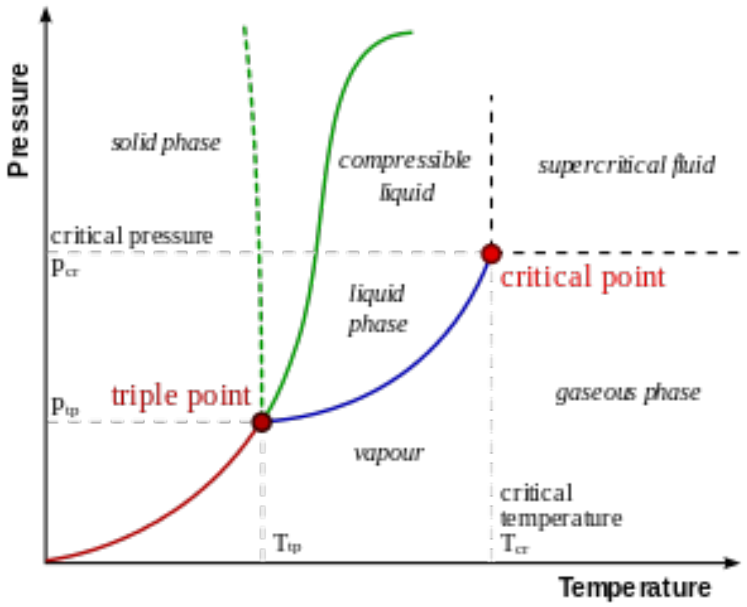

Figure 1.1: The phase diagram of water. The different phases of water are separated by the blue, red and green lines.

Usually different phases have different symmetries. The critical point is defined as the point where multiple phases co-exist but can not be distinguished because of identical properties (like volume, density and free energy). It is across this critical point that

phase transitions can be observed. Another example of phase transitions can be seen in magnets. These include the ferromagnetic phase transition in materials such as iron, where the magnetization, which is the first derivative of the free energy with respect to the applied magnetic field strength, increases continuously from zero as the temperature is lowered below the Curie temperature. The magnetic susceptibility, the second derivative of the free energy with the field, changes discontinuously. The phase with zero magnetization is called the paramagnetic phase while the other phase is ferromagnetic phase. There is clearly symmetry breaking while going from a paramagnetic phase to a ferromagnetic phase. We will be studying about a specific model of ferromagnetism called the Ising model.

1.1 The Ising Model

Ising model is a mathematical model of ferromagnetism in statistical mechanics. The model consists of spins arranged on a lattice in which each spin interacts with a specified set of spins, most usually with the nearest neighbours. The Hamiltonian of the simplest Ising model is given by

$H(\sigma) = - \sum_{<i,j>} J\sigma_i\sigma_j$

where σ is the set of spins and N is the number of spins. <i,j> stands for σ_i and σ_j being nearest neighbours (assuming external magnetic field is zero).

The difficulty of solving the above model (computing the partition function) depends on the dimensionality and the topology of the lattice. 1D Ising model is solvable exactly. It does not exhibit phase transitions whereas the 2D Ising model does. Phase transitions can only be seen in physical systems where N approaches ∞. Various techniques have been adopted to solve the Ising model like transfer matrix method, mean field theory and renormalization group theory.

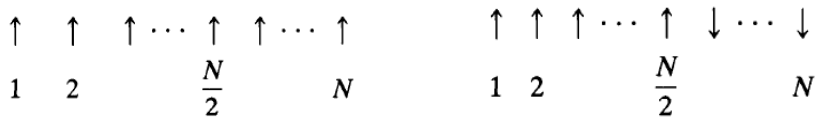

Figure 1.2: The diagram in the left is the ordered phase while the one on the left is a disordered phase with zero magnetization.

In 1D Ising model there is no phase transition possible as the formation of domain wall is favored at all temperatures as the energy cost is really low compared to the entropy change. At any non zero temperature formation of domain walls are favoured which is a paramagnetic phase. Hence the system prefers to stay in a paramagnetic phase at all temperatures, ruling out the possibility of phase transitions. The energy supplied to change an ordered phase to a disordered one is 2J which is well compensated by the entropy change that is proportional to $\log_e$N.

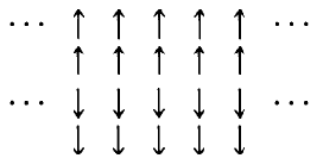

Figure 1.3: The disordered phase in a 2D Ising lattice with zero magnetization.

In the case of 2D Ising model, the excitation energy to go from an ordered state to a disordered one (like Fig. 1.3) goes as $\sqrt{N}$. At low temperatures this energy difference is sufficient to stabilize the ordered state. At high temperatures, the disordering is favored. Hence we can qualitatively see how the 2D Ising model can exhibit an order-disorder phase transition. For a d-dimensional Ising model, the excitation energy is $N^{\frac{d}{d+1}}$. Its also called the surface energy. Lars Onsagner solved the 2D Ising model exactly, where the critical temperature (T_c) was found to be $\frac{J}{0.44k_b}$

1.2 Mean Field Theory

An approximate yet simple approach to solve the 2D Ising model problem is mean field theory. This theory can qualitatively explain the theory of phase transitions. The idea is to treat each spin sitting independently in the mean field created by the other spins.

Consider a spin at lattice site **r**. The Hamiltonian for that particular spin is

$$H = -\beta\mathcal{H} = \sum_{\mathbf{r}'} K\sigma_r\sigma_{r'} = K\sigma_r \sum_{\mathbf{r}'} \sigma_{r'} \tag{1.1}$$

where **r'** is the set of nearest neighbours of **r**.

Mean field theory assumes that the fluctuating sum in Eq 1.1 can be replaced by the

sum of average values of the neighbouring spins. This means

$$H = K\sigma_r \sum_{\mathbf{r'}} \langle \sigma_{r'} \rangle = Kz\sigma_r \langle \sigma_r \rangle \tag{1.2}$$

where z is the number of nearest neighbours. The right side equality of Eq 1.2 is due to translational invariance of the Ising model.
Solving the above model, we obtain

$$\langle \sigma_r \rangle = tanh(Kz\langle \sigma_r \rangle) \tag{1.3}$$

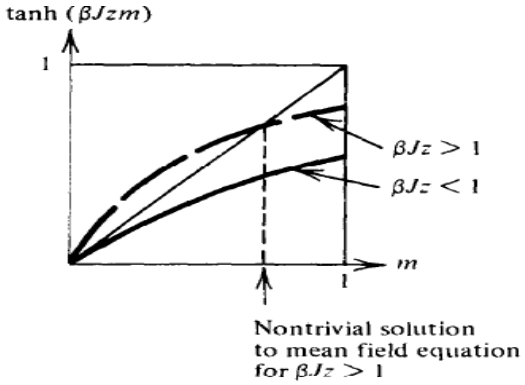

Figure 1.4: The mean field theory solutions for various values of coupling constants.

As we can see from fig 1.4 there is a non trivial solution if Kz>1 and for Kz<1 the only solution is m = 0. Hence we have arrived at the 2 phases , ferromagnetic (Kz > 1) and paramagnetic (Kz < 1) phase. When Kz = 1 , the phase transition occurs. According to MFT, the critical coupling constant = 1/z = 0.25.
Approximating Eq 1.3 around the point Kz = 1 using Taylor expansion,

$$\begin{aligned} m &= Kzm - \frac{Kzm^3}{3} \\ m^2 &= \frac{3(Kz-1)}{Kz} \\ m &= \sqrt{3\epsilon} \end{aligned} \tag{1.4}$$

where ϵ = (Kz - 1) = deviation from critical point.
This is the first time we encounter critical exponents. The other exponents can also be found by using Taylor expansion around the critical point.

Mean field theory was able to rightly predict the presence of a critical point and the

critical exponents even though they were numerically incorrect. This inaccuracy can be attributed to Eq. 1.2. Since z is just 4 in this case, this approximation can lead us to incorrect values. The higher z is, the more successful mean field theory is.

1.3 Monte Carlo Simulations on Ising Model

Since it is very hard to numerically calculate the physical properties of the 2D Ising model, we can run simulations to calculate these quantities to a high accuracy. We can use this data to find the critical point and the critical exponents to some extent. This can also be used as a tool to visualize the critical phenomena. The Metropolis Hastings algorithm was used to generate the samples below.

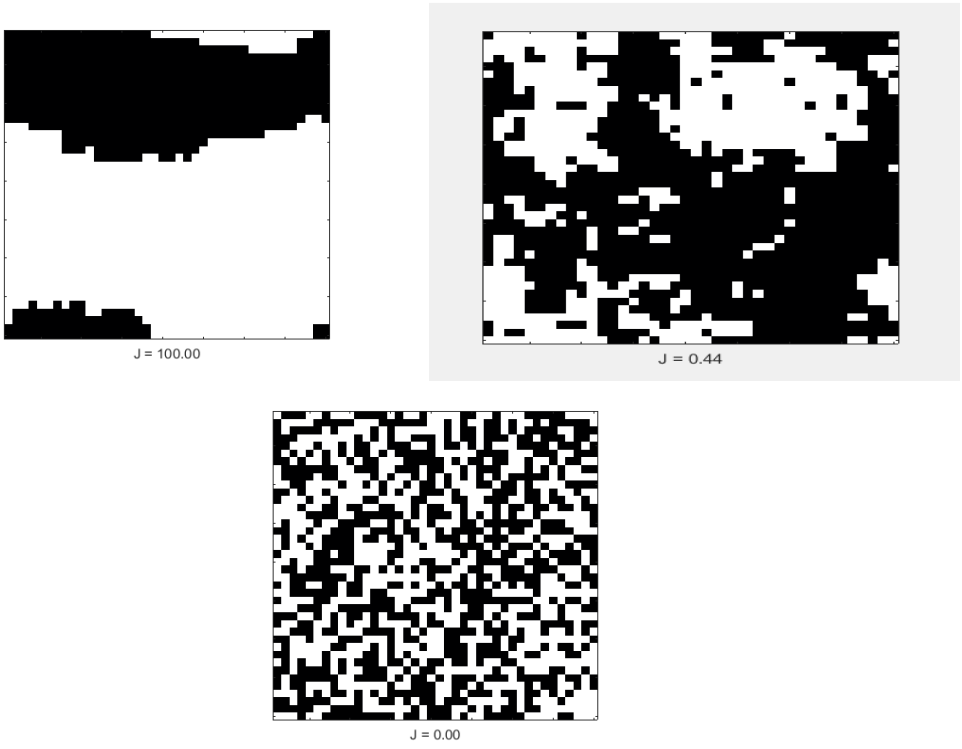

Figure 1.5: The samples obtained from the Monte Carlo simulation of 2D Ising model with 1600 spins for different coupling constants.

As we can observe at the critical point, we can see lot of relatively large clusters being formed. Also, we can observe large domain walls being formed even though the whole sample has zero magnetization. This collective behavior that has appeared just through the interaction with the neighbours is a very important feature of this model.

1.4 Criticality and Universality

Thermodynamic variables show singular or non-analytic behavior as one approaches the critical point. The singular behaviour in the vicinity of the critical point is characterized by a set of critical exponents. These sets of critical exponents have been experimentally observed to be the same in a variety of physical systems. For example, the liquid gas phase transition and the 2D Ising model have the same set of critical exponents even though they are different physical models. Physical systems with the same set of critical exponents are said to belong to a universality class. This, along with the scale invariance of physical systems near the critical point was an important motivation for physicist to use RG theory.

CHAPTER 2

Renormalization Group Theory

We previously saw that mean field theory was not able to predict the critical exponents accurately. Neither was there a framework to explain the universality. RG is an iterative coarse graining scheme that allows for the extraction of relevant features or operators as a physical system is examined at different length scales. This is done by integrating out the fluctuations at a lower length scale. In other words, we try to explain the properties of the original system using the new system which has lesser degrees of freedom. RG transformations must retain the symmetry of the original system. Ideally , the free energy must be invariant during each transformation. RG also explains the scale invariance of a system at the critical point. In this report we will be primarily discussing block spin renormalization procedures. The set of points we obtain while iteratively do RG transformations constitute the **RG flow**.

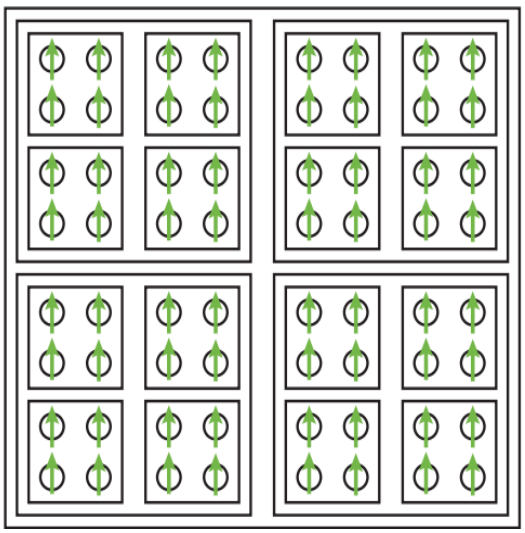

Figure 2.1: Here is an example of block spin renormalization. In the first iteration, spins are grouped by blocks of four and a single spin represents these spins in the new system. This procedure is repeated until we reach a critical point.

2.1 Basic Ideas of RG

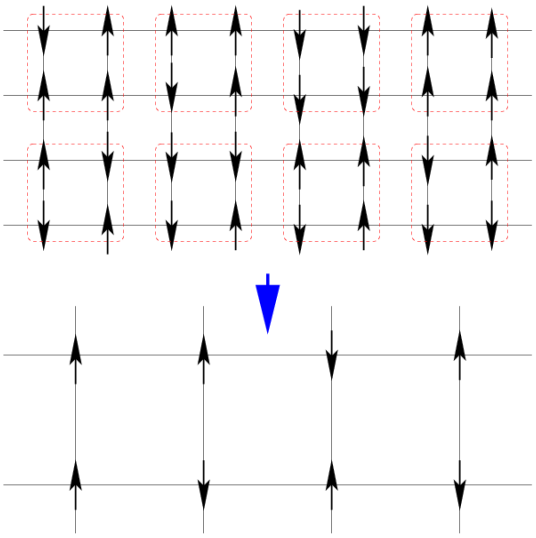

Figure 2.2: This corresponds to a single iteration of RG where each group of four spins are replaced by a single spin using the majority rule.

RG consists of two principal steps. The first step is constructing the coarse graining transformation where we define how the coupling constants transform. The second step is to find the critical point and the exponents using the defined transformation. Consider a system with Hamiltonian

$$H = -\beta\mathcal{H} = \sum_{\mathbf{n}} K_n X_n(S) \tag{2.1}$$

where K_n are the coupling constants and $X_n(S)$ are the operators which are functions of the degrees of freedom (S). a is the lattice spacing of the system with say N spins.

Let **K** represent the set of coupling constants and R^l be the RG transformation that groups together degrees of freedom in a block of linear dimension la.($l > 1$)

$$\begin{aligned} \mathbf{K}' &= R^{l_1}(\mathbf{K}) \\ \mathbf{K}'' = R^{l_2}(\mathbf{K}') &= R^{l_2}.R^{l_1}(\mathbf{K}) \\ \Rightarrow R^{l_1 l_2}(\mathbf{K}) &= R^{l_2}.R^{l_1}(\mathbf{K}) \end{aligned} \tag{2.2}$$

To find the fixed point , one has to either repeatedly do the RG procedure until it converges to the fixed point or solve the equation. The solution of Eq. 2.3 is called a **fixed**

point .

$$\mathbf{K} = R^l(\mathbf{K}) \tag{2.3}$$

There are many ways to construct the RG transformation for the given problem, but interestingly all these transformations should lead us to the same critical point and exponents.

2.2 Local Behaviour of RG Flows

In this section we will be studying how points near the fixed point transform under RG. Consider a point **K** near the fixed point **K*** such that

$$\mathbf{K} = \mathbf{K}^* + \delta\mathbf{K} \tag{2.4}$$

Performing an RG transformation on **K** , **K'** = R^l(**K**) = **K*** + δK'. To find δK', we use Taylor expansion. Writing component wise,

$$\mathbf{K'_n} = \mathbf{K^*_n} + \sum_m \frac{\partial K'_n}{\partial K_m}\Big|_{K_m=K_{m*}} .\delta K_m + O((\delta K)^2) \tag{2.5}$$

Let $M_{nm} = \frac{\partial K'_n}{\partial K_m}\Big|_{K_m=K_{m*}}$

$$\Rightarrow \delta K'_n = \sum_m M_{nm}.\delta K_m \tag{2.6}$$

Let the eigenvalues and eigenvectors of M^l be z$_l^\sigma$ and $\mathbf{e}^\sigma$ where σ labels the eigenvalues. From Eq. 2.2 , we can say that M$^{l_1}M^{l_2}$ = M$^{l_1 l_2}$

$$M^l\mathbf{e}^\sigma = z_\sigma^l\mathbf{e}^\sigma \tag{2.7}$$

$$\Rightarrow z_\sigma^{l_1} z_\sigma^{l_2} = z_\sigma^{l_1 l_2} \tag{2.8}$$

Solving the above functional equation, we get $z_\sigma^l = l^{y_\sigma}$.

The set y_σ are nothing but the critical exponents. Hence, just from the framework which consists of assumption in Eq. 2.2 , we are able to land at the critical exponents.

- If $y_\sigma > 0 \Rightarrow$ the field grows in the direction of $\mathbf{e}^\sigma$ after each RG transformation. It is also called a relevant eigen vector. The corresponding operator is called a relevant operator.
- If $y_\sigma < 0 \Rightarrow$ the field shrinks in the direction of $\mathbf{e}^\sigma$ after each RG transformation. It is called an irrelevant eigen vector. The corresponding operator is called a irrelevant operator.
- If $y_\sigma = 0 \Rightarrow$ the field does not in the direction of $\mathbf{e}^\sigma$ after each RG transformation. It is called a marginal eigen vector. The corresponding operator is called a marginal operator.

2.3 Solving Ising model using RG

2.3.1 1D Ising model

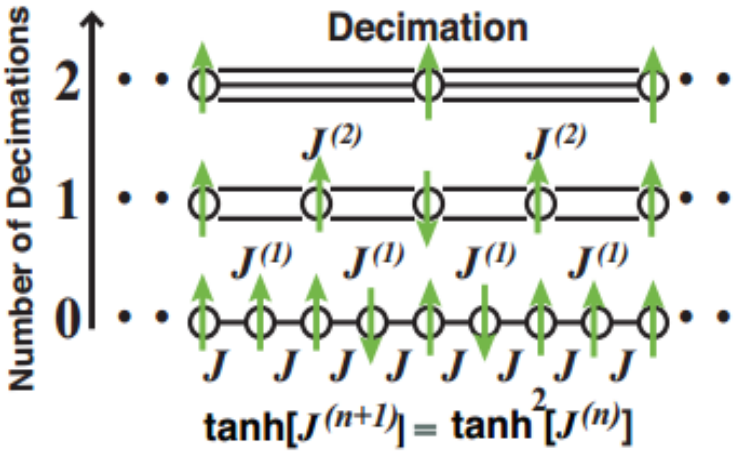

Figure 2.3: Decimation in 1D Ising model. This is done by integrating out the fluctuations due to the spins on the odd numbered lattice. This gives rise to a new coupling constant.

Ising model in one dimension can be solved exactly using RG. It should be no surprise as it is quite simple to compute the partition function. This specific technique we will be using is called decimation. We keep the spins at odd lattices unchanged while summing over the fluctuations due to spins at even lattices. Let Z(K,N) denote the partition function for this model with N spins and K as the coupling constant. Our aim is to write Z(K,N) as a function of the remain N/2 spins while transforming.

$$\begin{aligned} Z(K,N) &= \sum_{s_1,s_2,\ldots s_N} e^{K(s_1s_2+s_2s_3+\ldots+s_Ns_1)} \\ &= \sum_{s_1,s_3,s_5\ldots} (e^{K(s_1+s_3)} + e^{-K(s_1+s_3)})(e^{K(s_3+s_5)} + e^{-K(s_3+s_5)})\ldots \\ &= \sum_{s_1,s_3,s5\ldots} e^{K'(s_1s_3+s_3s_5+s_5s_7\ldots)} \\ &= Z(K', N/2)[f(K)]^{N/2} \end{aligned} \tag{2.9}$$

Solving for K' and f(K) in the above system we get, K' = $\frac{ln[cosh(2K)]}{2}$ and f(K) = 2 $\cosh^{1/2}$(2K)

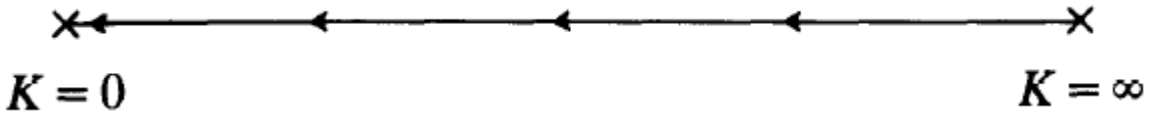

Figure 2.4: RG flow diagram for 1D Ising model.

K = 0 and K = ∞ are the fixed points. K = 0 corresponds to the totally disordered state while K = ∞ corresponds tot the totally ordered state. Notice when the system is viewed at K = 0 or ∞ it is independent of length scale.

2.3.2 2D Ising model

2D Ising model can be solved by using various RG schemes. We will talk about two of those briefly, decimation and bond moving. 2D Ising model is more interesting than the 1D case because there exists a non trivial point and we can use RG to find them.

Decimation in 2D Ising model is similar to that of the 1D case. But in this case the procedure can not be exactly carried out as decimating half of the spins give rise to more complicated couplings like 4-spin coupling, next nearest neighbour coupling and

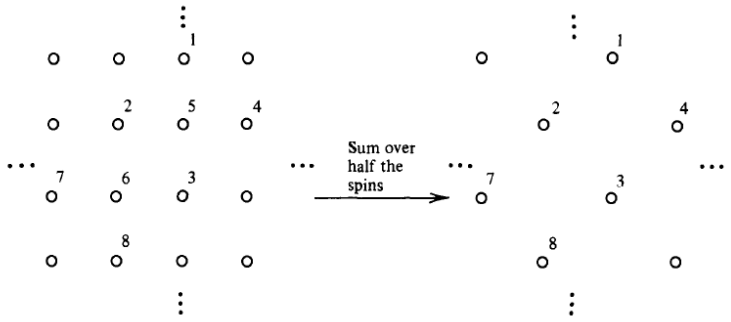

Figure 2.5: Decimation for 2D Ising model

so on. This is a tricky problem. If we try to include these couplings before decimation, we will be getting more complicated couplings. To deal with the problem, we neglect couplings that are weak and approximate the other couplings to fit the model.

$$Z(K,N) = \sum_{s_1,s_2,...s_N} e^{K(...+s_5(s_1+s_2+s_3+s_4)+s_6(s_2+s_3+s_7+s_8)+...)}$$

$$= \sum_{remains s_i} ...(e^{K(s_1+s_2+s_3+s_4)} + e^{-K(s_1+s_2+s_3+s_4)})(e^{K(s_2+s_3+s_7+s_8)} + e^{-K(s_2+s_3+s_7+s_8)})... \tag{2.10}$$

To see the effect of decimating s_5 , we write $e^{K(s_1+s_2+s_3+s_4)} + e^{-K(s_1+s_2+s_3+s_4)}$ in terms of s_1, s_2, s_3, s_4.

$$e^{K(s_1+s_2+s_3+s_4)} + e^{-K(s_1+s_2+s_3+s_4)} = f(K)[e^{K_1/2(s_1s_2+s_2s_3+s_3s_4+s_4s_1)+K_2(s_1s_3+s_2s_4)+K_3s_1s_2s_3s_4}] \tag{2.11}$$

Solving the above set of equations,

$$\begin{aligned} K_1 &= \frac{ln[cosh4K]}{4} \\ K_2 &= \frac{ln[cosh4K]}{8} \\ K_3 &= \frac{ln[cosh4K]}{8} - \frac{ln[cosh2K]}{2} \end{aligned} \tag{2.12}$$

Neglecting the K_3 coupling , and writing the couplings due to K_2 in terms of the

nearest neighbour coupling, we get the fixed point by solving

$$K = \frac{3ln[cosh4K]}{8} \tag{2.13}$$

We get K_c = 0.50 as the critical point which is 0.06 away from the actual value of 0.06.

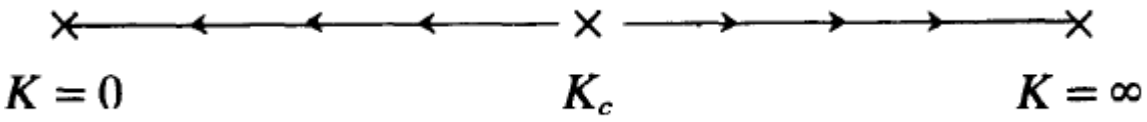

Figure 2.6: RG flow diagram for 2D Ising model.

K_c is unstable fixed point, as the flows move away from K_c

Notice the connectivity/topology of the lattice produces the long range order and phase transition. Once we start coarse graining, it leads to more complicated interactions. For example , when we integrated out s_5 , in the transformed model s_1 and s_3 interact due to the fluctuations of s_5.

2.3.3 Bond Moving Methods

This is a simple yet powerful method to solve the Ising model problem. In this method we add a potential to the Hamiltonian so that the renormalization becomes easier. Even though it is an approximate method, the results we get is quite close to the actual values.

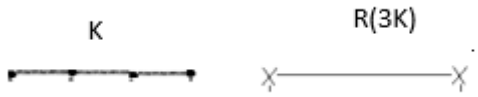

Figure 2.7: Decimating b spins to a single one for the special case of b = 3.

Consider b spins on a line(as shown in the Fig. 2.7). To renormalize the b bonds with a single bond, we have to renormalize the system to the scale of ba where a is the

lattice spacing. This is similar to the 1D decimation we saw earlier.

$$\begin{aligned} e^{K' s_A s_b + K'_0} &= \sum_{s_A, s_1, \ldots s_b} e^{K(s_1 s_2 + s_2 s_3 + \ldots + s_{b-1} s_b)} \\ \Rightarrow tanh K' &= (tanh K)^b \\ \Rightarrow K' = R^b(K) &= tanh^{-1}[(tanh K)^b] \end{aligned} \tag{2.14}$$

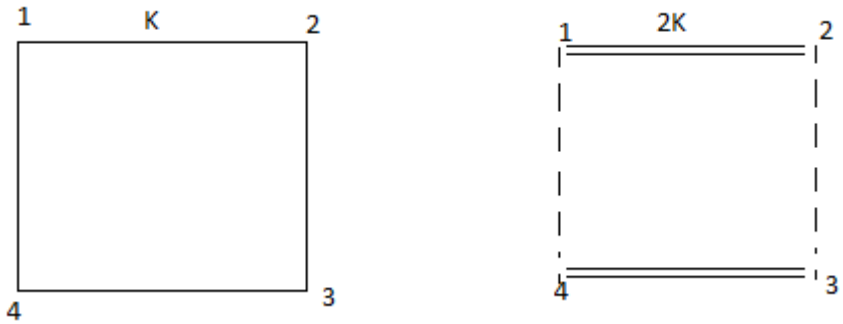

Figure 2.8: A simple case of bond moving.

Now, consider the 2D Ising model, with the standard nearest neighbour interaction. Now adding a potential of $e^{K(s_1 s_2 - s_2 s_3 + s_3 s_4 - s_4 s_1)}$ would imply that the new Hamiltonian would have terms like $s_2 s_3$ replaced by $s_1 s_2$. This is equivalent to moving the bond that connects s_2 and s_3 to s_1 and s_2.

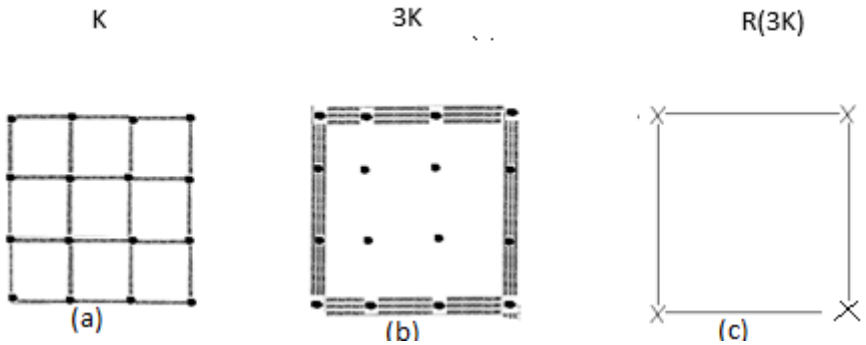

Figure 2.9: Bond moving for the 2D Ising model for the special case of $b = 3$.(b) is after the bond moving procedure. (c) is after renormalizing.

Extending this for the general case of renormalizing the 2D Ising model to a scale of ba , we get K' = R^b(bK). This can be extended to any real number b. To find the fixed point, we need to differentiate the above equation and set $\frac{\partial K'}{\partial b}\Big|_{b=1} = 0$. Differentiating

and setting b = 1 + l,

$$\begin{aligned} \frac{\partial K}{\partial l} &= K + (sinhK)(coshK)ln[tanhK] \\ \Rightarrow K* &= -(sinhK^*)(coshK^*)ln[tanhK^*] \end{aligned} \tag{2.15}$$

Solving 2.15, we get K* = 0.44 which is the exact value of the critical point in 2D Ising model.

Hence we were able to predict the fixed point accurately using the bond moving method. This can not be used to predict the critical exponents, since they involve derivatives of the partition function. Adding the potential will alter the derivative value, giving incorrect critical exponents. There are more sophisticated methods to calculate the exponents like variational RG methods which we would not be discussing here.

CHAPTER 3

Applications of RG in Deep Learning

Deep learning is a class of machine learning algorithms that use a cascade of many layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. The algorithms learn multiple levels of representations that correspond to different levels of abstraction, the levels form a hierarchy of concepts. There are lots of machines that implement deep learning. These algorithms have been very successful in recent times and there is very little understanding behind their success. In this chapter, we will be studying whether this success can be explained by RG with regard to Restricted Boltzmann Machines(RBMs).

3.1 Restricted Boltzmann Machines

Restricted Boltzmann Machine(RBM) is a generative artificial neural network that is used to model probability distribution. It is used for compressing data and for anomaly detection. RBMs have found applications in dimensionality reduction, classification, collaborative filtering, feature learning[5] and topic modelling. For example, in the domain of image processing, RBMs are used to reconstruct images from partially developed images. RBMs are used to model probability distribution from the training

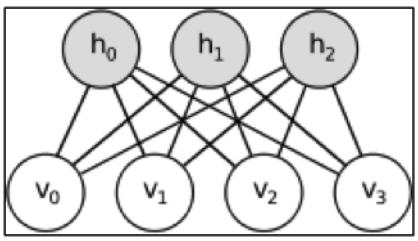

Figure 3.1: Restricted Boltzmann Machines.v_i are the visible units (input) where as h_i are the hidden nodes.

data. We will restrict our discussion to RBMs acting on binary data drawn from some probability distribution, P($\{v_i\}$), with $\{v_i\}$ binary spins labeled by an index i = 1...N.

For example, for black and white images each spin v_i encodes whether a given pixel is on or off and the distribution P($\{v_i\}$) encodes the statistical properties of the ensemble of images.

$$
\begin{aligned}
E(\{v_i\},\{h_j\}) &= \sum_j b_j h_j + \sum_{i,j} v_i W_{ij} h_j + \sum_i c_i v_i \\
p(\{v_i\},\{h_j\}) &= \frac{e^{-E(\{v_i\},\{h_j\})}}{\sum_{\{v_i\}}\sum_{\{h_j\}} e^{-E(\{v_i\},\{h_j\})}}
\end{aligned}
\tag{3.1}
$$

Here λ = { b_j, c_i, w_{ij} } are the parameters of the model. w_{ij} can be thought of how well the node i in the visible layer is correlated to node j in the hidden layer. $\{b_j\}$ and $\{c_i\}$ are the biases for each node.

$$
\begin{aligned}
p(\{v_i\}) &= \sum_{h_j} p(\{v_i\},\{h_j\}) \\
p(\{h_j\}) &= \sum_{v_i} p(\{v_i\},\{h_j\})
\end{aligned}
\tag{3.2}
$$

Our aim is to find the parameters (λ) so that $p_\lambda(\{v_i\})$ matches with the actual probability distribution(P($\{v_i\}$)) of the input. Since it is very hard to match the probability distribution, we can only get approximate it. This approximation can be done in several ways like minimising the Kullback-Leibler divergence between the true distribution of the data P($\{v_i\}$) and the variational distribution $p_\lambda(\{v_i\})$. The Kulback-Leibler divergence is defined as

$$
D_{KL}(P(\{v_i\})||p_\lambda(\{v_i\}) = \sum_{\{v_i\}} P(\{v_i\}) log \frac{P(\{v_i\})}{p_\lambda(\{v_i\})} \tag{3.3}
$$

If the RBM is able to exactly model the distribution, the divergence becomes zero.

For further reference we define quantities $H_\lambda^{RBM}[\{v_i\}]$ and $H_\lambda^{RBM}[\{h_j\}]$

$$
\begin{aligned}
p_\lambda(\{v_i\}) &= \frac{e^{-H_\lambda^{RBM}[\{v_i\}]}}{Z} \\
p_\lambda(\{h_j\}) &= \frac{e^{-H_\lambda^{RBM}[\{h_j\}]}}{Z}
\end{aligned}
\tag{3.4}
$$

3.2 Mapping RG to Deep Learning

RBM uses multiple layers of representation to automatically learn relevant features directly from structured data. The possible explanation of the working of RBM using RG can be attributed to the fact that both of them are trying to model p($\{v_i\}, \{h_j\}$) in their own ways. After each level of training , the relevant features continue to grow while the irrelevant features diminish similar to RG. RBMs assign a single node in the hidden layer for each sets of highly correlated visible nodes. Block spin renormalization does exactly the same by grouping spins that are close to each other into a single block. At the critical point, since the spins in the same block are highly correlated, it is reasonable to represent those spins by a single spin, which is the basis of block spin renormalization.

3.2.1 Variational Renormalization Group Theory

While studying RG, we came up with techniques like decimation to find the critical points and exponent. Variational RG introduces a new parameter $T_\lambda(\{v_i\}, \{h_j\})$ that models the interaction between the physical spins and the renormalized spins.Formally its defined as

$$e^{-H_\lambda^{RG}}[\{h_j\}] = \sum_{\{v_i\}} e^{T_\lambda(\{v_i\},\{h_j\})-H(\{v_i\})} \tag{3.5}$$

From Eq.(3.5) we can see that $T_\lambda(\{v_i\}, \{h_j\})$ models p($h_j|v_i$).In variational RG we choose the parameters λ so that the approximation error is the least, formally we would like to minimise the absolute $|F_\lambda^h - F_\lambda^v|$ where F^h is the free energy of the renormalized system.

In variational RG, the joint probability distribution is modelled by $T_\lambda(\{v_i\}, \{h_j\})$ and $E_\lambda(\{v_i\}, \{h_j\})$ in the RBMs. Hence it is safe to assume

$$\begin{aligned} e^{-E(\{v_i\},\{h_j\})} = p^{RBM}(\{v_i\}, \{h_j\}) = p^{RG}(\{v_i\}, \{h_j\}) = p^{RG}(\{h_j\}|\{v_i\})p^{RG}(\{v_i\}) \\ = e^{T_\lambda(\{v_i\},\{h_j\})-H[\{v_i\}]} \\ \Rightarrow T(\{v_i\}, \{h_j\}) = E(\{v_i\}, \{h_j\}) + H[\{v_i\}], \end{aligned} \tag{3.6}$$

Here $H[\{v_i\}]$ encodes the actual probability distribution $P[\{v_i\}]$. Given the above assumption, it is easy to prove that the Hamiltonian of the renormalized spins ($H^{RG}(\{h_j\})$) is the same as ($H^{RBM}(\{h_j\})$) defined in Eq. 3.4.

$$\frac{e^{-H_\lambda^{RG}}[\{h_j\}]}{Z} = \sum_{\{v_i\}} \frac{e^{T_\lambda(\{v_i\},\{h_j\})-H(\{v_i\})}}{Z} = \sum_{\{v_i\}} \frac{e^{E(\{v_i\},\{h_j\})}}{Z} = \sum_{\{v_i\}} \frac{p^{RBM}((\{v_i\},\{h_j\}))}{Z}$$
$$= p^{RBM}(\{h_j\}) = \frac{e^{-H_\lambda^{RBM}}[\{h_j\}]}{Z}$$
$$\Rightarrow H_\lambda^{RG}[\{h_j\}] = H_\lambda^{RBM}[\{h_j\}] \tag{3.7}$$

3.2.2 Results

We used RBMs stacked on top of each other as shown in Fig. 3.2 with the output of a layer being the input of the next layer.

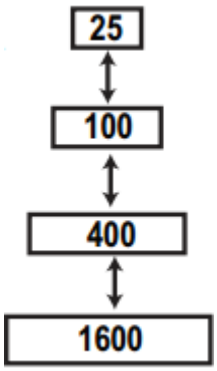

Figure 3.2: A Deep RBM with four layers of size 1600, 400, 100, and 25 spins

The RBM was trained with samples generated using the Monte Carlo simulation near the critical point on a 40X40 lattice. Since the correlation length diverges near the critical point, the spins in a block are expected to be highly correlated. This can be observed in the correlation diagram(Fig. 3.3), where the correlation between each spin in the deep layer and a spin in the visible layer is plotted.

From Fig.(3.3) we can see that the RBM is implementing a coarse graining procedure similar to the block spin renormalization. It is also observed that the blocks observed at each level is approximately of the same size. This local block spin structure emerges from the training process, suggesting this deep RBM is self-organizing to

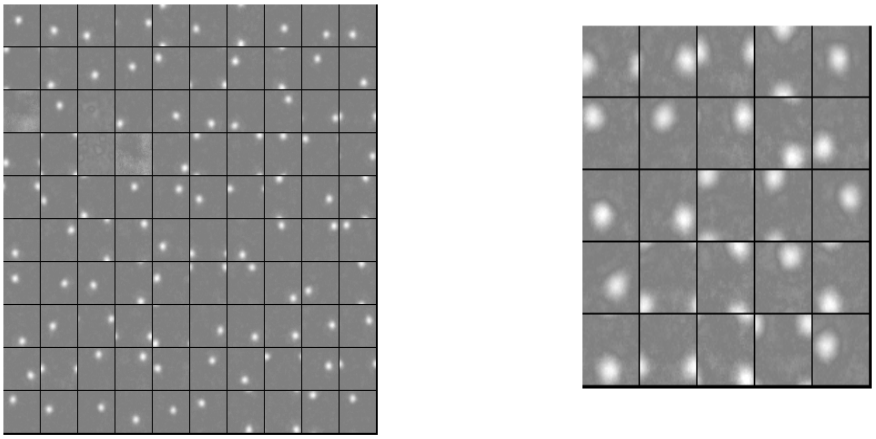

Figure 3.3: The figure on the left shows the correlation diagram for the layer with 100 spins and the other one is the correlation diagram for the layer with 25 spins. Each 40X40 pixel represents a spin in the hidden layer. Gray stands for low correlation and white stands for high correlation.

implement block spin renormalization.

3.3 Conclusion and Further Work

It is very interesting to observe RBMs implement something similar to block spin renormalization while learning samples of 2D Ising model near critical point. But this mapping of RBM to RG has been done for a very specific case. We can work on a more mathematically strong approach to establish the equivalence. If we are able to model the working of RBMs using operators, so that after each application of an RG transformation, they either grow (relevant operators) or diminish (irrelevant operators), we will be able to have a more concrete understanding of RBM's connection to RG theory. Once this equivalence is established, we can use it to solve problems in one domain using the techniques from the other domain.

REFERENCES

1. Introduction to Modern Statistical Mechanics by David Chandler
2. Lectures on Phase Transitions and the Renormalization Group by Nigel Goldenfeld
3. Real Space Renormalization edited by Burkhardt, Theodre, Leeuwen, J.M.J. van (Eds.)
4. Statics, Dynamics and Renormalization by L.P.Kadanoff
5. An exact mapping between the Variational Renormalization Group and Deep Learning by Pankaj Mehta
6. Learning Deep Architectures for AI by Yoshua Bengio