The Kac-Ward approach to the Ising Model

Lis, M.

2014

Introduction

1.1 Historical background

The Ising model, introduced by Lenz [42] in 1920 and named after his student Ising, is one of the simplest models of statistical physics exhibiting a phase transition. It was proposed as a model for spontaneous magnetization in ferromagnets. This phenomenon occurs when, in the absence of an external magnetic field, a piece of ferromagnetic material, e.g. iron, becomes magnetized whenever its temperature drops below a certain critical level (called the Curie temperature). Ising [35] proved that the one dimensional model does not account for the existence of this phenomenon and asserted that the same should hold in higher dimensions. More than a decade later his conclusion was proved wrong by Peierls [48] who, with the use of a geometric argument, established that in dimensions higher than one the Ising model does exhibit spontaneous magnetization at low temperatures. This discovery gave a positive answer to the important question whether a model of statistical physics can explain the existence of phase transitions in natural processes.

The critical point, i.e. the inverse temperature at which the phase transition occurs, for the Ising model defined on the square lattice was first identified by Kramers and Wannier [41] as the fixed point of a certain duality transformation. The first rigorous proof of criticality of the self-dual point was given by Onsager [47], who explicitly computed the free energy density of the model. The planar Ising model with no external field has been since known as one of the few exactly solvable models of statistical physics.

In search for a solution simpler than the famously complicated algebraic method of Onsager, Kac and Ward [36] proposed a combinatorial approach, where the partition function of the Ising model is given by the determinant of a certain geometrically defined matrix. However, their arguments were of heuristic nature and a key assumption turned out to be incorrect (see Section 1.3). At the same time when the technical details required for a rigorous account of this method were considered in the physics literature, Fisher [25] and Kasteleyn [39] independently discovered the dimer approach. This new method allowed to rederive the result of Onsager by means of yet another combinatorial determinant.

Over the years, many different approaches have been developed to answer various questions concerning the Ising model. Even though the model has a simple formulation, the mathematical landscape surrounding it has become vast and diverse. We choose to mention the recent discrete holomorphic approach developed by Smirnov et al. (see the references in Chapter 5). It allowed to finally prove many results on the scaling limits of the Ising model which were predicted in the 1980s within the framework of conformal field theory. As we will show in Chapter 5, notions crucial for the discrete holomorphic approach have very natural interpretations in terms of the matrix introduced by Kac and Ward.

1.2 The Ising model and its phase transition

Let $\mathbb{Z}^2$ be the nearest neighbour square lattice (see Figure 1.1). The vertices of $\mathbb{Z}^2$ represent molecules in a piece of ferromagnetic material and the edges indicate which pairs of molecules interact via their magnetic moments. Consider a bounded region of the plane and let $\mathcal{G} = (V, E)$ be the finite subgraph of $\mathbb{Z}^2$ induced by the molecules contained in this region. Let $\Omega = \{-1, +1\}^V$ be the space of spin configurations, which consists of all possible assignments of the two opposite magnetic moments to the molecules. The energy of a configuration $\sigma \in \Omega$ with no external magnetic field is given by the Hamiltonian

$$\mathscr{H}(\sigma) = -\sum_{uv\in E} J_{uv}\sigma_u\sigma_v,$$

where the numbers $J_{uv} > 0$, called coupling constants, express the strength of interaction between the molecules u and v. Note that in the case of $\mathbb{Z}^2$, one usually considers the homogenous model with all coupling constants equal to 1. One defines the Ising model on $\mathcal{G}$ at inverse temperature $\beta > 0$ as the probability measure on Ω, which is proportional to the Gibbs-Boltzmann weights, i.e.

$$\mathbf{P}_{\mathcal{G},\beta}(\sigma) \propto \exp(-\beta\mathscr{H}(\sigma)).$$

Note that since the coupling constants are ferromagnetic (i.e. positive), the model assigns larger probability to configurations where the spins are aligned with their neighbours. The partition function $\mathcal{Z}_{\mathcal{G},\beta}$ of the model is the inverse proportionality constant which makes the Gibss-Boltzmann weights into a probability measure, i.e.

$$\mathcal{Z}_{\mathcal{G},\beta} = \sum_{\sigma\in\Omega} \exp\big(-\beta\mathscr{H}(\sigma)\big).$$

If the Ising model is an accurate simplification of the physical process occurring in ferromagnets, we should be able to observe a phase transition, i.e. a drastic change of the properties of the model when we increase β through the critical value β_c corresponding to the Curie temperature. This change would be expressed in some form of non-analytic

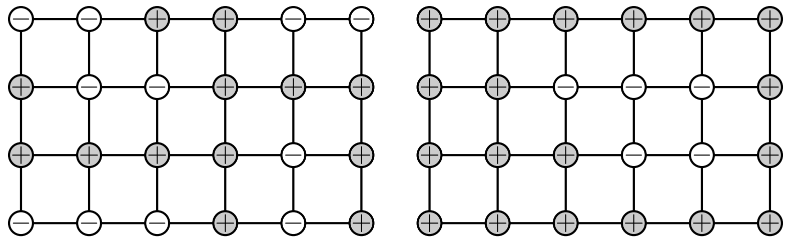

Figure 1.1: Two spin configurations with free (left) and positive (right) boundary conditions on a rectangular piece of the square lattice $\mathbb{Z}^2$.

behavior at β_c of the physical quantities describing the model. However, we defined the Ising model only in bounded volumes, and such quantities, like the free energy density

$$f_{\mathcal{G}}(\beta) = -\frac{1}{\beta|V|} \ln \mathcal{Z}_{\mathcal{G},\beta},$$

are in this case always analytic in $\beta \in (0, \infty)$, as they are defined in terms of finite sums of analytic functions. This is why a phase transition can become visible only in the thermodynamic limit; when we look at larger and larger regions of the plane, and as a result, $\mathcal{G}$ increases to $\mathbb{Z}^2$. The formalism of statistical physics predicts that the critical point of a physical system should then correspond to a singular point of the thermodynamic limit of some physical quantity. An important example is the infinite-volume free energy density

$$f(\beta) := \lim_{\mathcal{G} \to \mathbb{Z}^2} f_{\mathcal{G}}(\beta)$$

of the homogenous Ising model. Onsager [47] was the first to explicitly compute f and prove that it does indeed posses a singular point (see Corollary 2.4 for the exact formula). More precisely, he showed that the specific heat of the system, which is expressed in terms of second-order derivatives of f with respect to temperature (see e.g. [3]), diverges at the self-dual point of Kramers and Wannier. This established that β_c and the self-dual point are indeed the same. Onsager also proved that the first derivative of f is continuous, and hence the Ising model is said to exhibit a second-order phase transition.

Investigating the analytic properties of the free energy density is just one way of looking at the phenomenon of phase transition. Another approach, and perhaps a more natural one in the setting of the Ising model, is to look at the behavior of spins. One of the simplest ways to do this is to analyze the so called one-point function, which is the expected value of a single spin:

$$\langle \sigma_v \rangle^+_{\mathcal{G},\beta} = \sum_{\sigma \in \Omega} \sigma_v \mathbf{P}^+_{\mathcal{G},\beta}(\sigma), \qquad v \in V,$$

where $\mathbf{P}^{+}_{\mathcal{G},\beta}$ is the Ising probability measure with positive boundary conditions, i.e. the Ising measure conditioned on the boundary spins having value $+1$ (see Figure 1.1). A positive or negative one-point function indicates the preference of a spin to be positive or negative respectively, and a vanishing one-point function indicates no preference for a sign. Due to the effect of positive boundary conditions, $\langle \sigma_v \rangle^{+}_{\mathcal{G},\beta}$ is strictly positive for all β (this follows from the standard high-temperature expansion, see Section 2.4.1). In other words, in finite volume, the presence of positive boundary spins causes the bulk spins to prefer the $+1$ state at all temperatures. However, in the thermodynamic limit, the boundary moves further and further away, and for temperatures high enough, its influence on a fixed particle vanishes. To be precise, let

$$\langle \sigma_v \rangle^{+}_{\mathbb{Z}^2,\beta} := \lim_{\mathcal{G}\to\mathbb{Z}^2} \langle \sigma_v \rangle^{+}_{\mathcal{G},\beta}$$

be the infinite-volume one-point function, called spontaneous magnetization. The exact formula for the spontaneous magnetization of the homogenous Ising model was announced by Onsager in 1949 and its first derivation was given three years later by Yang [57] (see [5] for the history of this result). From this analytic solution, it follows in particular that the spontaneous magnetization is strictly positive for $\beta \in (\beta_c, \infty)$, and is equal to zero for $\beta \in (0, \beta_c]$. This contrasting behavior agrees with the physical picture of ferromagnetism and shows that β_c is critical also in terms of the behavior of spins.

There are also other relevant quantities whose properties change at the critical point. In this text, we will analyze the two-point functions

$$\langle \sigma_u \sigma_v \rangle^{\square}_{\mathcal{G},\beta} = \sum_{\sigma\in\Omega} \sigma_u \sigma_v \mathbf{P}^{\square}_{\mathcal{G},\beta}(\sigma), \quad \langle \sigma_u \sigma_v \rangle^{\square}_{\mathbb{Z}^2,\beta} := \lim_{\mathcal{G}\to\mathbb{Z}^2} \langle \sigma_u \sigma_v \rangle^{\square}_{\mathcal{G},\beta} \qquad u, v \in V,$$

where $\square \in \{+, \mathrm{free}\}$ denotes the imposed boundary conditions. The Ising measure with free boundary conditions $\mathbf{P}^{\mathrm{free}}_{\mathcal{G},\beta}$ is the unconditional measure $\mathbf{P}_{\mathcal{G},\beta}$ which does not put any restrictions on the values of boundary spins. As we will see, the contrasting behavior in this context is that, when u and v goes to infinity, $\langle \sigma_u \sigma_v \rangle^{\square}_{\mathbb{Z}^2,\beta}$ converges exponentially fast to zero for $\beta \in (0, \beta_c)$, or stays bounded away from zero for $\beta \in (\beta_c, \infty)$. One then says that the system is in the disordered or ordered state respectively.

1.3 The approach of Kac and Ward

Let $\mathcal{G} = (V, E)$ be a finite planar graph and let $x = (x_e)_{e\in E}$ be a vector of complex weights on the edges of $\mathcal{G}$. One defines the transition matrix indexed by the directed edges of $\mathcal{G}$ by

$$\Lambda_{\overrightarrow{uv},\overrightarrow{wz}}(x) = \begin{cases} x_{uv} e^{i\angle(\overrightarrow{uv},\overrightarrow{wz})/2} & \text{if } v = w \text{ and } u \neq z; \\ 0 & \text{otherwise,} \end{cases} \tag{1.1}$$

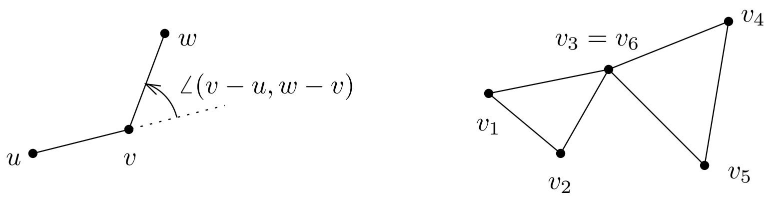

Figure 1.2: The turning angle from the vector $v - u$ to the vector $w - v$ (left). The loop $(v_1, v_2, v_3, v_4, v_5, v_6)$ on the right has sign -1, the loop $(v_1, v_2, v_3, v_5, v_4, v_6)$ has sign $+1$.

where uv is the undirected version of $\vec{uv}$, and $\angle(\vec{uv}, \vec{vz})$ is the turning angle $\angle(v - u, w - v) \in (-\pi, \pi]$ from $v - u$ to $w - v$ (see Figure 1.2). Here, we identify the vertices with the complex numbers. The Kac–Ward matrix (operator) is defined as

$$T(x) = \mathrm{Id} - \Lambda(x), \tag{1.2}$$

where Id is the identity matrix.

Kac and Ward argued in a heuristic way that the square of the partition function of the Ising model on a subgraph of $\mathbb{Z}^2$ is proportional to the determinant of the above matrix[1] with an appropriately chosen weight vector. One of their key "theorems", as pointed out by Sherman [50], turned out to be actually false. Sherman proposed another approach based on a conjecture of Feynman which involved infinite products over certain classes of closed paths in the graph. However, his arguments were also incomplete [51] and the manipulations of infinite products non-rigorous. A simplification of Sherman's method was presented by Burgoyne [11] but it also lacked mathematical rigour.

In [55], Vdovichenko presented a different way of approaching the problem. We will now briefly summarize it. The starting point is expressing the Kac–Ward determinant as a sum over configurations of loops in the graph. To this end, one uses the formula[2]

$$\det^{\frac{1}{2}}\big(\mathrm{Id} - \Lambda(x)\big) = \exp\bigg(-\sum_{r=1}^{\infty} \mathrm{tr}\Lambda^r(x)/2r\bigg), \tag{1.3}$$

which is valid whenever the spectral radius of $\Lambda(x)$ is strictly smaller than one. To prove this identity, it is enough to represent the determinant and trace in terms of the eigenvalues of $\Lambda(x)$ and use the power series of the logarithm around 1. One then

[1] Actually, Kac and Ward considered the matrix $\mathrm{Id} + \Lambda(x)$. In the case of $\mathbb{Z}^2$, this matrix has the same determinant as $\mathrm{Id} - \Lambda(x)$ (this follows from (1.3) and the fact that $\mathbb{Z}^2$ admits only loops of even length). The general definition (1.2) appeared for the first time in the article of Dolbilin et al. [21].

[2] Note that the same identity with the discrete Laplacian in place of the Kac-Ward operator forms the basis for the random walk representation of the Gaussian free field [10].

interprets $-\mathrm{tr}\Lambda^r(x)/2r$ as a weighted sum over all loops of length r in the graph. Note that the transition matrix (1.1) assigns zero weight to a backtracking step, and hence the paths and loops appearing in the Kac-Ward method are always non-backtracking. Since the winding angle of a loop is a multiple of 2π, the complex factors containing the turning angle in (1.1) always multiply to ± 1 in the weight of the loop. This is why we sometimes refer to this method as the signed loop approach. By a result of Whitney [56], when a loop does not go through any edge more than once, its sign can be interpreted in terms of the number of times the loop crosses itself (see Figure 1.2, and Chapter 2 for precise definitions). By expanding the exponential, one can finally write the determinant as a weighted sum over configurations of loops. The crucial step then is to prove that the determinant yields the partition function of the Ising model. To this end, one has to prove that the weights of configurations with repeated edges sum up to zero. This is done by constructing a bijection on the space of configurations of loops, which flips the signs of the configurations while preserving the remaining factors of the weights (see Chapter 2 for a detailed account).

As pointed out in [20], Vdovichenko made a combinatorial mistake by not taking into account what we call in this text the multiplicity of a loop (2.7). The first mathematically rigorous account of the Kac–Ward method seems to be given by Dolbilin et al. [21] over thirty years after Vdovichenko's paper. To prove the relationship between the partition function of the Ising model and the determinant of the Kac–Ward matrix, the authors compared the coefficients of these two quantities treated as polynomials in appropriate variables x. Hence, they circumvented the problems of analyzing infinite generating functions.

The interest in the Kac–Ward method has been recently revived. Cimasoni generalised the approach to arbitrary surface graphs and considered relations between critical Kac–Ward matrices, Kasteleyn matrices, Laplacians and discrete holomorphicity [15–17]. A complete account of the signed-loop approach, together with the derivation of the critical point of the Ising model on the square lattice, was given by Kager, Meester and the author [38]. Cimasoni and Duminil-Copin [18] used the Kac–Ward determinants to compute the critical temperature of Ising models defined on doubly periodic graphs. Helmuth [31] put the signed-loop method into a general combinatorial framework of heaps of pieces. In [44, 45], the author analyzed phase transitions in Ising models on general planar graphs and established a link between the Kac–Ward approach and the discrete holomorphic approach of Smirnov et al.

1.4 Overview

Chapter 2 is based on the article

[38] KAGER W., LIS M. AND MEESTER R. The Signed Loop Approach to the Ising model: Foundations and Critical Point. *J. Stat. Phys.* 152, 2 (2013), 353–387.

In this chapter, we explore the foundations of the Kac–Ward method à la Vdovichenko, including details that have so far been neglected or overlooked in the literature. We provide complete, rigorous and detailed proofs of the combinatorial identities central to the signed loop method, all in a geometric manner. We essentially follow the same steps as in Vdovichenko's paper and we show how to fix her error by including a loop's multiplicity into its weight, as we do in equation (2.7) below.

In addition to clarifying the signed loop approach, we also apply the results to the Ising model on $\mathbb{Z}^2$ in ways not considered before to derive both new and classical results about the Ising model. More precisely, we demonstrate how to obtain explicit formal expressions for the free energy density and two-point functions in terms of sums over loops. By our bound on the spectral radius of the transition matrix on the square lattice, these expressions are valid all the way up to the self-dual point. As a result, we rigorously obtain Onsager's formula for the free energy density and we show the contrasting behavior of the two-point functions below and above the self-dual point. As a corollary, it follows that the self-dual point is critical both for the analytic properties of the free energy density, and for the decay of the two-point functions.

Chapter 3 is based on unpublished work.

In this chapter, we fill the "gap" of Chapter 2, which consists in only considering the positive boundary conditions in the low-temperature regime and the free boundary conditions in the high-temperature regime. To this end, we introduce additional combinatorial objects which are related to the notions from Chapter 2, and we analyze their properties. We also briefly discuss how to handle the multi-point functions, i.e. correlation functions with more than two spins. As a side result, we rederive an interesting theorem of Boel, Groeneveld and Kasteleyn [7] about correlation functions of boundary spins (see Theorem 3.3).

Chapter 4 is based on the article

[45] LIS, M. Phase transition free regions in the Ising model via the Kac-Ward operator. *Comm. Math. Phys.* (2014).

In this chapter, we provide an upper bound on the spectral radius of the Kac–Ward transition matrix on a general planar graph. Combined with the Kac–Ward formula for the partition function of the planar Ising model, this allows us to identify regions in the complex plane where the free energy density limits are analytic functions of the inverse temperature. The bound turns out to be optimal in the case of isoradial graphs, i.e. it yields criticality of the self-dual Z-invariant coupling constants introduced by Baxter [4]. The class of self-dual Z-invariant Ising models contains, for example, the critical homogeneous models on the square, triangular and hexagonal lattice. To the best of our knowledge, our result is the first of this type which also applies to aperiodic graphs.

We also briefly discuss how, by using the methods of Chapters 2 and 3, we can conclude that the self-dual Z-invariant Ising models are critical in terms of correlation functions.

Chapter 5 is based on the article

[44] LIS, M. The Fermionic Observable in the Ising Model and the Inverse Kac–Ward Operator. *Ann. Henri Poincaré* (2013).

In this chapter, we show that the critical Kac–Ward operator on isoradial graphs acts in a certain sense as the operator of s-holomorphicity. This notion of strong discrete holomorphicity was introduced by Smirnov [54] and is central to the discrete holomorphic approach to the Ising model. We prove that a function is s-holomorphic if and only if it lies in the kernel of the critical Kac–Ward operator composed with some natural projection operator. Moreover, we prove that the inverse Kac–Ward operator can be identified with the Green's function of a discrete Riemann–Hilbert boundary value problem similar to the ones appearing in [13, 14, 32, 34, 53].

Furthermore, using bounds obtained in Chapter 4 for the spectral radius and operator norm of the Kac–Ward transition matrix, we provide a general picture of the non-backtracking walk representation of the critical and supercritical (high-temperature) inverse Kac–Ward operators on isoradial graphs. As a consequence, the solution to the discrete Riemann–Hilbert boundary value problem looks similar to the random walk represention of the solution to the discrete Dirichlet boundary value problem for harmonic functions. The important difference is that the non-backtracking walks have complex weights which do not yield a probability measure on the space of walks.

Summary

The main contributions of this thesis are:

+ providing a rigourous account of the Kac–Ward approach à la Vdovichenko and showing how it can be applied to describe the phase transition in the Ising model,

+ identifying non-critical values of the inverse temperature in a way which applies to any planar graph and yields criticality of the self-dual Z-invariant Ising models on isoradial graphs (and in particular, of the critical homogenous models on the square, triangular and hexagonal lattice),

+ showing how the notions of the discrete holomorphic framework of Smirnov et al. arise within the Kac–Ward approach, and providing a non-backtracking walk representation of the solution to a relevant Riemann–Hilbert boundary value problem, which may be of independent interest.