VU Research Portal

The Kac-Ward approach to the Ising Model

Lis, M.

2014

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
Lis, M. (2014). *The Kac-Ward approach to the Ising Model*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 13. Mar. 2024

Foundations and the critical point on the square lattice

2.1 Introduction and main results

In this chapter, we consider the Ising model on finite rectangles in $\mathbb{Z}^2$, by which we mean graphs $\mathcal{G} = (V, E)$ whose vertex sets V consist of all points of $\mathbb{Z}^2$ contained in a rectangle $[a, b] \times [c, d]$ in $\mathbb{R}^2$, and whose edge sets E consist of all unordered pairs $\{u, v\}$ with $u, v \in V$ such that their L_1 distance $\|u - v\|$ is 1. For brevity, we will also write uv instead of $\{u, v\}$ for the edge between u and v (note that $uv = vu$). The *boundary* $\partial\mathcal{G}$ of $\mathcal{G}$ is the set of those vertices u in V for which there is a $v \in \mathbb{Z}^2 \setminus V$ with $\|u - v\| = 1$.

We associate to $\mathcal{G}$ a space of *spin configurations* $\Omega_{\mathcal{G}} = \{-1, +1\}^V$. For $\sigma \in \Omega_{\mathcal{G}}$, σ_v denotes the spin at v. Sometimes we impose *positive boundary conditions*, meaning that we restrict ourselves to the set of configurations

$$\Omega^+_{\mathcal{G}} = \{\sigma \in \Omega_{\mathcal{G}} \colon \sigma_u = +1 \text{ if } u \in \partial\mathcal{G}\}.$$

In contrast, when we speak of *free boundary conditions*, we work with the unrestricted set of spin configurations $\Omega^{\text{free}}_{\mathcal{G}} = \Omega_{\mathcal{G}}$.

The Ising model defines a (Gibbs–Boltzmann) probability distribution on the set of spin configurations. At inverse temperature β, it is given by

$$\mathbf{P}^{\square}_{\mathcal{G},\beta}(\sigma) = \frac{1}{\mathcal{Z}^{\square}_{\mathcal{G},\beta}} \prod_{uv \in E} e^{\beta\sigma_u\sigma_v}, \qquad \sigma \in \Omega^{\square}_{\mathcal{G}}, \tag{2.1}$$

where $\square \in \{\text{free}, +\}$ stands for the imposed boundary condition, and $\mathcal{Z}^{\square}_{\mathcal{G},\beta}$ is the *partition function* of the model, defined as

$$\mathcal{Z}^{\square}_{\mathcal{G},\beta} = \sum_{\sigma \in \Omega^{\square}_{\mathcal{G}}} \prod_{uv \in E} e^{\beta\sigma_u\sigma_v}. \tag{2.2}$$

To simplify the notation, and following the physics literature, we will write

$$\langle \cdot \rangle^{\square}_{\mathcal{G},\beta} = E^{\square}_{\mathcal{G},\beta}(\cdot) \tag{2.3}$$

for expectations with respect to $\mathbf{P}^{\square}_{\mathcal{G},\beta}$. Important functions are the *free energy density*

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

and its *infinite-volume* limit

$$f(\beta) = \lim_{\mathcal{G}\to\mathbb{Z}^2} f^{\square}_{\mathcal{G}}(\beta).$$

It is well known that this limit exists, and also not difficult to show that the limit is the same for all boundary conditions, see for example [52, Section II.3] (as we shall see, existence of the limit actually also follows from the signed loop approach for non-critical β). The formalism of statistical mechanics predicts that phase transitions coincide with discontinuities or other singularities in derivatives of the free energy density.

In the case of the Ising model on $\mathbb{Z}^2$, the classical arguments of Peierls [48] and Fisher [26] (see also [19, 28, 29]) established that it does undergo a phase transition, which can be characterized in terms of a change in behaviour of the infinite-volume limits $\langle\sigma_u\sigma_v\rangle^{\square}_{\mathbb{Z}^2,\beta}$ of the *two-point functions* $\langle\sigma_u\sigma_v\rangle^{\square}_{\mathcal{G},\beta}$. Peierls' argument implies that as $\|u-v\|\to\infty$, these infinite-volume two-point functions stay bounded away from 0 for large enough β. In contrast, Fisher's argument yields exponential decay to 0 of these two-point functions as $\|u-v\|\to\infty$, for sufficiently small β. However, there is a gap between the two ranges of β for which these arguments work, so they are not strong enough to conclude that the phase transition is sharp. As we shall see, the signed loop method studied here does lead to a proof of this fact.

To be more specific, a signed loop is essentially a closed, non-backtracking walk, with a positive or negative weight assigned to it; see Section 2.1.1 for a precise definition. In this chapter, we demonstrate how the free energy and two-point functions can be expressed as (infinite) sums over signed loops in $\mathbb{Z}^2$, and moreover, how the rate of convergence of these sums can be controlled. For instance, Theorem 2.3 below expresses the free energy density $f(\beta)$ as an explicit formal sum of the weights of all loops for which the origin is the smallest vertex visited in lexicographic order. Likewise, Theorems 2.5 and 2.7 express the two-point functions $\langle\sigma_u\sigma_v\rangle^{\square}_{\mathbb{Z}^2,\beta}$ as infinite sums over explicitly defined classes of signed loops.

In all cases, we first derive the corresponding loop expressions for finite rectangles $\mathcal{G}$. These were also obtained simultaneously and independently by Helmuth [31] via the theory of heaps of pieces, but we in addition give explicit bounds on the rates of convergence to take the infinite-volume limit. As corollaries, without requiring any external results, we rederive several classical results about the Ising model, which can be summarized as follows:

Corollary 2.1 (Sharpness of phase transition). *The free energy density $f(\beta)$ is analytic for all $\beta > 0$ except at the self-dual point $\beta = \beta_c$ given by*

$$\exp(-2\beta_c) = \tanh\beta_c = \sqrt{2}-1. \tag{2.4}$$

Moreover, as $\|u-v\|\to\infty$, $\langle\sigma_u\sigma_v\rangle^{\text{free}}_{\mathbb{Z}^2,\beta}$ decays to 0 exponentially fast when $\beta\in(0,\beta_c)$, while $\langle\sigma_u\sigma_v\rangle^{+}_{\mathbb{Z}^2,\beta}$ stays bounded away from 0 when $\beta\in(\beta_c,\infty)$.

This chapter is organized as follows. A precise definition of signed loops and the formulation of our main results follow in Sections 2.1.1–2.1.4, with our results for the Ising model in Section 2.1.2, and our key combinatorial identities in Section 2.1.4. The proofs of these identities are given in Section 2.2, and the proofs of our results about the Ising model are in Sections 2.4.2, 2.4.3 and 2.4.4.

2.1.1 Signed loops

Although in this chapter our applications are in $\mathbb{Z}^2$, it will be necessary to study signed loops on a more general class of graphs. Our starting point is a (finite or infinite) graph $\mathcal{G} = (V, E)$ embedded in the plane, with vertex set V and edge set E. We identify $\mathcal{G}$ with its embedding. We assume $\mathcal{G}$ does not have multiple edges, but we do not assume that $\mathcal{G}$ is planar. For convenience (although this is not strictly necessary), we require that edges are straight line segments in the embedding, and that except for the vertices at the two endpoints, no other vertices lie on an edge. As before, we write uv or vu for the (undirected) edge between u and v.

A *walk* of n *steps* in $\mathcal{G}$ is a sequence $(v_0, v_1, \ldots, v_{n-1}) \in V^n$ such that $v_i v_{i+1} \in E$ for $i = 0, 1, \ldots, n-2$, and $v_{i+2} \neq v_i$ for $i = 0, 1, \ldots, n-3$ (walks are *non-backtracking*). If all rotations (i.e. circular shifts) of the sequence $(v_0, \ldots, v_{n-1})$ are also walks (so that in particular, $v_0 v_{n-1} \in E$), then we call the walk *closed.* We now order the vertices of $\mathcal{G}$ lexicographically by their coordinates in the plane, and define a *loop* as a closed walk $(v_0, \ldots, v_{n-1})$ which is the lexicographically smallest element in the collection consisting of all rotations of $(v_0, \ldots, v_{n-1})$ and all rotations of the reverse sequence $(v_{n-1}, \ldots, v_0)$ (note that these are all in a way closed walks traversing the same loop).

If $\ell = (v_0, \ldots, v_{n-1})$ is a loop or a closed walk, we shall make the identification $v_j \equiv v_{j \bmod n}$ for all $j \in \mathbb{Z}$. We say that a loop ℓ is *edge-disjoint* if $v_i v_{i+1} \neq v_j v_{j+1}$ for all $i, j \in \{0, \ldots, n-1\}$ with $i \neq j$. If ℓ is not edge-disjoint, it might be the case that the sequence $(v_0, \ldots, v_{n-1})$ is periodic, in which case we call ℓ a *periodic loop.* The *multiplicity* of ℓ, denoted by $m(\ell)$, is its number of steps divided by its smallest period. In particular, the multiplicity of every nonperiodic loop is 1.

Given two distinct edges uv and vw, we define $\angle(v - u, w - v) \in (-\pi, \pi)$ as the turning angle in the plane from the vector $v - u$ to $w - v$, see Figure 1.2 (left). The *winding angle* $\alpha(\ell)$ of a loop $\ell = (v_0, \ldots, v_{n-1})$ is simply the sum of all turning angles along the loop, that is,

$$\alpha(\ell) = \sum_{i=0}^{n-1} \angle(v_{i+1} - v_i, v_{i+2} - v_{i+1}). \tag{2.5}$$

We now define the *sign* $\operatorname{sgn}(\ell)$ of ℓ as

$$\operatorname{sgn}(\ell) = -\exp\Big(\frac{i}{2}\alpha(\ell)\Big). \tag{2.6}$$

Observe that the winding angle of every loop is a multiple of 2π (here we use the fact that the edges of $\mathcal{G}$ are straight line segments), hence the sign of a loop is either $+1$ or -1.

To define the *signed weight* of a loop, we require a vector $x = (x_{uv})_{uv \in E}$ of *edge weights* $x_{uv} \in \mathbb{R}$ (or $\mathbb{C}$). Given these edge weights x_{uv}, the signed weight of a loop $\ell = (v_0, \dots, v_{n-1})$ in $\mathcal{G}$ is defined as

$$w(\ell; x) = \frac{\operatorname{sgn}(\ell)}{m(\ell)} \prod_{i=0}^{n-1} x_{v_i v_{i+1}}. \tag{2.7}$$

Remark 2.2. *If a loop is edge-disjoint, it follows from Whitney's formula [56] that the sign of the loop is* -1 *if and only if the loop crosses itself an odd number of times (see Figure 1.2). For loops that are not edge-disjoint, it may not be so clear what is meant by a "crossing", but definition* (2.6) *makes sense for both kinds of loop. However, if one draws loops in such a way that each visit to an edge is drawn slightly apart from a previous visit, the number of crossings one is forced to draw will always be odd for a loop of sign* -1*, and even for a loop of sign* $+1$ *(see for instance Figure 2.4 below).*

2.1.2 Main results for the Ising model

We now return to the Ising model on $\mathbb{Z}^2$. In this section, we will formulate the main theorems for the Ising model, which express the free energy density and two-point functions in terms of sums over signed loops. Each of our theorems will be accompanied by a corollary, which taken together, constitute Corollary 2.1. Similar results as the ones presented here can be obtained for the hexagonal and triangular lattices using the same methods. In fact, the method applies to the Ising model on even more general (isoradial) graphs (see Chapter 4), and also allows one to study general multi-point functions (see Chapter 3).

We start with the free energy density $f(\beta)$. As we shall prove, $f(\beta)$ can be expressed as a sum over those loops in $\mathbb{Z}^2$ for which the origin $o = (0,0)$ is the lexicographicallyically smallest vertex traversed. To be more specific, we define $\mathcal{L}_r^\circ(\mathbb{Z}^2)$ as the collection of all loops $\ell = (v_0, \dots, v_{r-1})$ of r steps in $\mathbb{Z}^2$ such that $v_0 = o$. We take all edges of $\mathbb{Z}^2$ to have the same edge weight x. The weights $w(\ell; x)$ of all loops $\ell \in \mathcal{L}_r^\circ(\mathbb{Z}^2)$ are now defined by (2.7), where by slight abuse of notation, we let x denote both the weight of a single edge, and the vector of all edge weights. Write

$$f_r^\circ(x) = \sum_{\ell \in \mathcal{L}_r^\circ(\mathbb{Z}^2)} w(\ell; x).$$

Theorem 2.3. *The free energy density satisfies*

$$-\beta f(\beta) = \begin{cases} \ln(2\cosh^2\beta) + \sum_{r=1}^\infty f_r^\circ\big(\tanh\beta\big) & \text{if } \beta \in (0, \beta_c), \\ 2\beta + \sum_{r=1}^\infty f_r^\circ\big(\exp(-2\beta)\big) & \text{if } \beta \in (\beta_c, \infty). \end{cases} \tag{2.8}$$

Note that since $f_r^\circ(x) = f_r^\circ(1)\, x^r$, $\sum_r f_r^\circ(x)$ is really a power series in the variable x. The power series expressions (2.8) show directly that the free energy density is an analytic function of β on $(0, \beta_c) \cup (\beta_c, \infty)$. Thus, in terms of the behaviour of the free energy density, the Ising model can only be critical at the self-dual point β_c. That $f(\beta)$ is not analytic at β_c follows from Onsager's formula, which we will obtain as a corollary to Theorem 2.3:

Corollary 2.4 (Onsager's formula). *For $\beta \in (0, \beta_c)$ and for $\beta \in (\beta_c, \infty)$, the free energy density $f(\beta)$ is given by the formula*

$$-\frac{1}{\beta}\frac{1}{8\pi^2}\int_0^{2\pi}\int_0^{2\pi} \ln\bigl[4\cosh^2 2\beta - 4\sinh 2\beta(\cos\omega_1 + \cos\omega_2)\bigr]\, d\omega_1\, d\omega_2.$$

The functions f and $\partial(\beta f)/\partial\beta$, which is the *internal energy density* of the system, are both continuous functions of β. However, in [47] Onsager proved that the *specific heat*, which is the derivative of the internal energy density with respect to temperature, diverges as β approaches β_c. This shows that β_c is indeed critical for the behaviour of the free energy.

Next, we look at the magnetic behaviour of the model by considering the one-point and two-point functions above and below β_c. We start with the case $\beta \in (\beta_c, \infty)$. What we will show is that for fixed $u, v \in \mathbb{Z}^2$, the functions $\langle \sigma_u \rangle^+_{\mathcal{G},\beta}$ and $\langle \sigma_u \sigma_v \rangle^+_{\mathcal{G},\beta}$ have infinite-volume limits along rectangles $\mathcal{G}$, where the limits can be identified in terms of sums over certain classes of loops in $\mathbb{Z}^{2*}$, the dual graph of $\mathbb{Z}^2$, defined as follows. Given $u, v \in \mathbb{Z}^2$, let γ be a self-avoiding walk in $\mathbb{Z}^2$ from u to v (see Figure 2.1, left). We call a loop in $\mathbb{Z}^{2*}$ *uv-odd* if it crosses γ an odd number of times. Similarly, we call a loop in $\mathbb{Z}^{2*}$ *u-odd* if it crosses a self-avoiding walk γ in $\mathbb{Z}^2$ from u to ∞ an odd number of times. It is not difficult to see that neither of these definitions depends on the particular choice of γ. We write $\mathcal{L}_r^u(\mathbb{Z}^{2*})$ and $\mathcal{L}_r^{uv}(\mathbb{Z}^{2*})$ for the sets of u-odd and uv-odd loops in $\mathbb{Z}^{2*}$ of r steps, respectively.

Let x be the constant vector of edge weights on $\mathbb{Z}^{2*}$ equal to $\exp(-2\beta)$ for every edge, and define the weights of loops in $\mathbb{Z}^{2*}$ by (2.7). Set

$$f_r^u(x) = \sum_{\ell \in \mathcal{L}_r^u(Z^{2*})} w(\ell; x); \qquad f_r^{uv}(x) = \sum_{\ell \in \mathcal{L}_r^{uv}(\mathbb{Z}^{2*})} w(\ell; x).$$

Theorem 2.5. *For all $\beta \in (\beta_c, \infty)$ and $u, v \in \mathbb{Z}^2$ ($u \neq v$),*

$$\lim_{\mathcal{G}\to\mathbb{Z}^2} \langle \sigma_u \rangle^+_{\mathcal{G},\beta} = \exp\biggl(-2\sum_{r=1}^{\infty} f_r^u(x)\biggr) =: \langle \sigma_u \rangle^+_{\mathbb{Z}^2,\beta} > 0;$$

$$\lim_{\mathcal{G}\to\mathbb{Z}^2} \langle \sigma_u \sigma_v \rangle^+_{\mathcal{G},\beta} = \exp\biggl(-2\sum_{r=1}^{\infty} f_r^{uv}(x)\biggr) =: \langle \sigma_u \sigma_v \rangle^+_{\mathbb{Z}^2,\beta} > 0.$$

As a corollary to the proof of this theorem, we will obtain that for $\beta \in (\beta_c, \infty)$, the two-point functions in the infinite-volume limit stay bounded away from 0 when $\|u - v\| \to \infty$:

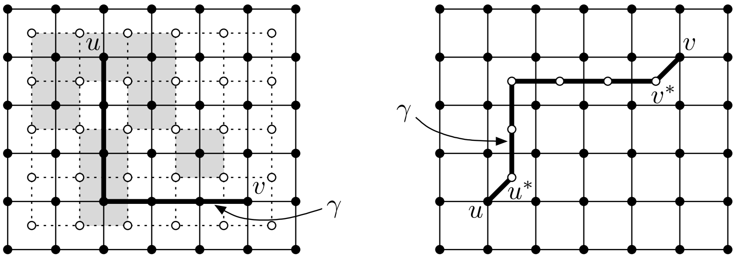

Figure 2.1: The walks γ (with bold edges), that we use to study the 2-point functions $\langle\sigma_u\sigma_v\rangle^{\square}_{\mathcal{G},\beta}$. The low-temperature case is on the left (spins are -1 in the gray squares, $+1$ in the white regions), the high-temperature case on the right.

Corollary 2.6 (Positive two-point functions above β_c). *For all $\beta \in (\beta_c, \infty)$,*

$$\lim_{\|u-v\|\to\infty} \langle\sigma_u\sigma_v\rangle^{+}_{\mathbb{Z}^2,\beta} = \left[\langle\sigma_o\rangle^{+}_{\mathbb{Z}^2,\beta}\right]^2 > 0.$$

We now turn to the two-point functions for $\beta \in (0, \beta_c)$. Fix $u, v \in \mathbb{Z}^2$, and choose dual vertices u^* and v^* of $\mathbb{Z}^{2*}$ such that $\|u - u^*\| = \|v - v^*\| = 1$. This choice is not unique, but every choice of u^* and v^* will do. Next, choose a self-avoiding walk γ in $\mathbb{Z}^{2*}$ from u^* to v^* (see Figure 2.1, right). Let V_γ denote the set of vertices in γ, and let E_γ denote the union of $\{uu^*, vv^*\}$ with the set of edges traversed by γ. Write $\mathbb{Z}^2_\gamma$ for the graph obtained from $\mathbb{Z}^2$ by adding the vertices and edges from V_γ and E_γ to it.

We define $\mathcal{L}^{uu^*}_r(\mathbb{Z}^2_\gamma)$ as the collection of loops in $\mathbb{Z}^2_\gamma$ that visit the edge uu^* exactly once and have $r - |E_\gamma|$ steps. Note that by definition, if $\ell \in \mathcal{L}^{uu^*}_r(\mathbb{Z}^2_\gamma)$, r only counts the number of steps taken by ℓ along edges of $\mathbb{Z}^2$, that is, the steps along the edges in E_γ are excluded. As our edge weight vector on $\mathbb{Z}^2_\gamma$ we take the vector x'_γ such that the weight of every edge in E_γ is 1, the weight of every edge in $\mathbb{Z}^2$ which intersects γ is $-\tanh\beta$, and the weight of all other edges is $\tanh\beta$. Set

$$f^{uu^*}_r(x'_\gamma) = \sum_{\ell\in\mathcal{L}^{uu^*}_r(\mathbb{Z}^2_\gamma)} w(\ell; x'_\gamma), \tag{2.9}$$

with $w(\ell; x'_\gamma)$ defined by (2.7). Let β^* denote the inverse temperature which is dual to β, i.e. such that $\exp(-2\beta^*) = \tanh\beta$.

Theorem 2.7. *For all $\beta \in (0, \beta_c)$ and $u, v \in \mathbb{Z}^2$ ($u \neq v$),*

$$\lim_{\mathcal{G}\to\mathbb{Z}^2} \langle\sigma_u\sigma_v\rangle^{\text{free}}_{\mathcal{G},\beta} = \left(\sum_{r=1}^{\infty} f^{uu^*}_r(x'_\gamma)\right)\langle\sigma_{u^*}\sigma_{v^*}\rangle^{+}_{\mathbb{Z}^{2*},\beta^*} =: \langle\sigma_u\sigma_v\rangle^{\text{free}}_{\mathbb{Z}^2,\beta}.$$

The term $\langle\sigma_{u^*}\sigma_{v^*}\rangle^{+}_{\mathbb{Z}^{2*},\beta^*}$ appearing here is the infinite-volume limit of a two-point function for an Ising model on the dual square lattice $\mathbb{Z}^{2*}$ at the dual inverse temperature β^* with positive boundary conditions. It can be expressed in terms of signed loops by means of Theorem 2.5. As an aside, we note that the result in Theorem 2.7 simplifies when u and v are on the same face of $\mathbb{Z}^2$ (i.e. $\|u-v\|=1$), since then we can take $u^*=v^*$, so that $\sigma_{u^*}\sigma_{v^*}=1$. Moreover, since the walk γ is void in this case, none of the edge weights will be equal to $-\tanh\beta$.

As a corollary to Theorem 2.7 we will obtain that the two-point functions decay exponentially to 0 with the distance $\|u-v\|$ for $\beta\in(0,\beta_c)$, which together with Corollaries 2.4 and 2.6 implies Corollary 2.1:

Corollary 2.8 (Decaying two-point functions below β_c). *For all $\beta\in(0,\beta_c)$ and all $u,v\in\mathbb{Z}^2$, we have that*

$$0\le\langle\sigma_u\sigma_v\rangle^{\text{free}}_{\mathbb{Z}^2,\beta}\le 16\sum_{r\ge\|u-v\|}\Big(\frac{\tanh\beta}{\tanh\beta_c}\Big)^r.$$

Note that Corollaries 2.6 and 2.8 show contrasting behaviour of the two-point functions above and below β_c: at low temperatures they stay bounded away from 0, while for high temperatures they decay to 0. However, we have used different boundary conditions above and below the critical point. In Chapter 3, we will remove this asymmetry from our results and we will complete the picture of the phase transition of the Ising model on the square lattice.

2.1.3 Additional edges and loop length

The set of edges E_γ that we introduced above to formulate our Theorem 2.7 is an example of what we call *additional edges*. As this example shows, we occasionally need these additional edges in our applications. They act as “shortcuts” that our loops can follow, and in general, just as we did above, we do not want to count the steps taken by our loops along these shortcuts.

Another example of the use of additional edges is in our proofs leading to Corollary 2.4, in which we compare loops in $\mathbb{Z}^2$ with loops on a torus. Here we face a problem, because our methods and theorems about signed loops (to be presented below) require that the graph we work on is embedded in the plane. As a solution, we will not work on the torus directly, but on a representation of it in the plane. As our representation, we take a rectangle in $\mathbb{Z}^2$ with opposite sides connected by additional edges, as illustrated in Figure 2.2. In this example, the additional edges do not correspond to edges that can be traversed by a loop on the torus, and this is the reason why steps taken along the additional edges again should not be counted.

In general, these considerations lead us to allow the edge set E of the graph $\mathcal{G}=(V,E)$ we work on to be divided into a set E_A of additional edges and a set $E\setminus E_A$ of edges that we call *representative*. For reasons that will become clear, we must impose that the set E_A is such that the graph (V,E_A) is free of cycles, but otherwise, the edge set can in principle be any subset of edges. We now define the *length* $r(\ell)$ of a loop $\ell=(v_0,\dots,v_{n-1})$ as the number of i in $\{0,\dots,n-1\}$ such that $v_iv_{i+1}\in E\setminus E_A$. Note the distinction between the *length* of a loop and its *number of steps*.

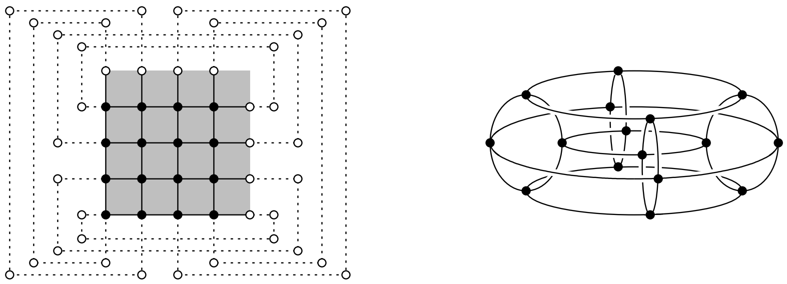

Figure 2.2: A square lattice wrapped around a torus (right) and a representation of it in the plane (left). The gray square corresponds to the torus, the dotted lines and open circles are the additional edges and vertices.

2.1.4 The combinatorial identities

We next formulate our combinatorial identities about signed loops for a fixed finite graph $\mathcal{G} = (V, E)$ embedded in the plane, satisfying the same assumptions as in Section 2.1.1. In particular, recall that edges are straight line segments, and that we do not assume $\mathcal{G}$ is planar, which implies that two edges can intersect in a point which is not a vertex. In this case, we say that the two edges *cross* each other.

We call a subset F of E *even* if every vertex in the subgraph (V, F) of $\mathcal{G}$ has even degree (the empty set is also even). By $C(F)$ we denote the total number of unordered pairs of edges in F that cross each other. Given a vector $x = (x_{uv})_{uv \in E}$ of edge weights on $\mathcal{G}$, we now define the *generating function $Z(x)$ of even subgraphs* of $\mathcal{G}$ as

$$Z(x) = \sum_{\text{even } F \subset E} (-1)^{C(F)} \prod_{e \in F} x_e. \tag{2.10}$$

If the graph $\mathcal{G}$ is planar, we can embed it in such a way that no edges cross each other, so that $C(F) = 0$ for all F, but in general, an even $F \subset E$ may give a negative contribution to the right-hand side of (2.10). This makes our generating function different from the one usually studied in the literature. Note as a consequence that different embeddings of the same (abstract) graph can lead to different functions $Z(x)$. Since we identify $\mathcal{G}$ with its embedding, this last fact does not concern us here.

Our combinatorial identities express the generating function $Z(x)$ in terms of sums over the signed loops in $\mathcal{G}$, with their weights defined by (2.7). This is what allows us to study the Ising model in terms of signed loops, since the free energy and the two-point functions of the Ising model can be expressed in terms of graph generating functions, as we shall see.

Theorem 2.9. *Let Δ be the maximum degree of the graph $\mathcal{G}$. If $|x_e| < (\Delta - 1)^{-1}$ for all $e \in E$, then*

$$Z(x) = \exp\biggl(\sum_{\ell \text{ in } \mathcal{G}} w(\ell; x) \biggr). \tag{2.11}$$

We will show that under the condition of Theorem 2.9, the loop weights are absolutely summable, so that the order of summation does not matter. In particular, let $\mathcal{L}_r$ be the collection of all loops of length r in $\mathcal{G}$, and let

$$f_r(x) = \sum_{\ell \in \mathcal{L}_r} w(\ell; x). \tag{2.12}$$

Then Theorem 2.9 implies that $Z(x)$ equals $\exp\bigl(\sum_r f_r(x)\bigr)$, but we claim that this latter equality already holds under a significantly weaker condition.

This condition can be formulated in terms of the *transition matrix* $\Lambda(x)$, which we now introduce. If uv is an edge of $\mathcal{G}$, then by $\overrightarrow{uv}$ we will denote the *directed* edge from u to v. The matrix $\Lambda(x)$ will be indexed by the directed *representative* edges of $\mathcal{G}$. Given two directed representative edges $\overrightarrow{uv}$ and $\overrightarrow{wz}$, we say that v is *linked to* w if either $v = w$, or there exists a sequence of distinct *additional* edges $v_1v_2, v_2v_3, \dots, v_{n-1}v_n$ such that $v = v_1$ and $v_n = w$. In the former case, if $v = w$ and $u \neq z$, we write

$$\angle(\overrightarrow{uv}, \overrightarrow{wz}) = \angle(v - u, z - w)$$

for the turning angle from $\overrightarrow{uv}$ to $\overrightarrow{wz}$. In the latter case, the sequence $(v_1, \dots, v_n)$ is a walk (the *chain*) linking v to w, passing through additional edges only, and we say that "$v \rightsquigarrow w$ via $(v_1, \dots, v_n)$". By our assumption that the additional edges form no cycles, there can be at most one such walk. Hence, without ambiguity, if v is linked to w in this way, we can define

$$\begin{aligned}\angle(\overrightarrow{uv}, \overrightarrow{wz}) &= \angle(v - u, v_2 - v) \\ &\qquad + \sum_{i=1}^{n-2} \angle(v_{i+1} - v_i, v_{i+2} - v_{i+1}) + \angle(w - v_{n-1}, z - w).\end{aligned}$$

The transition matrix $\Lambda(x)$ is now defined as follows. Write $\Lambda_{\overrightarrow{uv},\overrightarrow{wz}}(x)$ for the entry of the matrix with row index $\overrightarrow{uv}$ and column index $\overrightarrow{wz}$. Then

$$\Lambda_{\overrightarrow{uv},\overrightarrow{wz}}(x) = \begin{cases} x_{uv} e^{i\angle(\overrightarrow{uv},\overrightarrow{wz})/2} & \text{if } v = w \text{ and } u \neq z; \\ x_{uv} \prod_{i=1}^{n-1} x_{v_i v_{i+1}} e^{i\angle(\overrightarrow{uv},\overrightarrow{wz})/2} & \text{if } v \rightsquigarrow w \text{ via } (v_1, \dots, v_n); \\ 0 & \text{otherwise.} \end{cases} \tag{2.13}$$

Let $\lambda_i(x)$, $i = 1, 2, \dots, 2|E \setminus E_A|$, denote the eigenvalues of $\Lambda(x)$, and let $\rho(x) = \max_i |\lambda_i(x)|$ be its spectral radius. We will show that if $\rho(x) < 1$, then the $f_r(x)$ are absolutely summable. This leads to our second identity, which forms the core of the signed loop approach:

Theorem 2.10. *If* $\rho(x) < 1$*, then*

$$Z(x) = \exp\Bigl(\sum_{r=1}^{\infty} f_r(x)\Bigr) = \det^{\frac{1}{2}}\bigl(\mathrm{I} - \Lambda(x)\bigr). \tag{2.14}$$

Clearly, to apply Theorem 2.10 to the Ising model on $\mathbb{Z}^2$, we will need a bound on the spectral radius $\rho(x)$. Since $\rho(x)$ is bounded from above by the operator norm $\|\Lambda(x)\|$ of $\Lambda(x)$ induced by the Euclidean norm, the desired bound is provided by the next theorem:

Theorem 2.11. *For a finite rectangle $\mathcal{G}$ in $\mathbb{Z}^2$ with no additional edges,*

$$\begin{aligned} \|\Lambda(x)\| &\leq (\sqrt{2}+1)\|x\|_\infty, \\ \textit{and} \quad |f_r(x)| &\leq 2|V|r^{-1}(\sqrt{2}+1)^r\|x\|_\infty^r. \end{aligned}$$

If we take all edge weights to be 1, Theorem 2.11 says that the "number" of signed loops of n steps, counted with signs and multiplicities included, grows (in absolute value) like $(\sqrt{2}+1)^n$. Contrast this with the number of unsigned non-backtracking loops in $\mathbb{Z}^2$, which grows like 3^n. It is this reduction in growth rate which allows us to go all the way to the critical point, while the classical Peierls and Fisher arguments stay far from it. Indeed, in Section 2.1.2 we have seen that we will take our edge weights to be either $\exp(-2\beta)$ or $\tanh\beta$, so by Theorem 2.11 and since $\exp(-2\beta_c) = \tanh\beta_c = (\sqrt{2}+1)^{-1}$ (2.4), the spectral radius will be smaller than 1 for all $\beta \in (\beta_c, \infty)$ or all $\beta \in (0, \beta_c)$, respectively.

We conclude this introduction with a few remarks about the history and status of the combinatorial identities presented above. As already mentioned in the introduction, Kac and Ward observed in their paper [36] that the Onsager–Kaufman formula for the partition function $\mathcal{Z}_{\mathcal{G},\beta}$ of the Ising model on $\mathbb{Z}^2$ appears to be proportional to the square root of the determinant of a matrix, which for rectangles in $\mathbb{Z}^2$ is equivalent to our matrix $\mathrm{I} - \Lambda(x)$. Various attempts were subsequently undertaken to justify the formula $\mathcal{Z}_{\mathcal{G},\beta}^2 \propto \det T_{\mathcal{G}}$, and to rederive the Onsager–Kaufman formula in this way. These attempts involved expanding the partition function into a formal infinite product over signed loops [11, 50, 51], or a formal infinite sum over signed loop configurations [55]. In either case, the correct interpretation and convergence of the obtained formal expressions are serious mathematical issues.

These issues were circumvented by Dolbilin et al. [21] by directly comparing the coefficients of the finite polynomials $\mathcal{Z}_{\mathcal{G},\beta}^2$ and $\det T_{\mathcal{G}}$, thus rigorously proving the formula $\mathcal{Z}_{\mathcal{G},\beta}^2 \propto \det T_{\mathcal{G}}$ for finite planar graphs $\mathcal{G}$. The same method was then employed by Cimasoni to generalize this Kac–Ward formula to graphs embedded in surfaces of higher genus [15]. He also exposed a direct relation between the Kac–Ward determinant and the adjacency matrix arising in the dimer approach.

Historically then, the main focus appears to have been on the equality between the extreme left-hand and right-hand sides of equation (2.14), in cases where $Z(x)$ is proportional to the Ising partition function and $T_{\mathcal{G}} = \mathrm{I} - \Lambda(x)$. For the applications to the Ising model presented in this paper, however, the first equality in (2.14) is the more important and relevant one. Therefore, our proof proceeds along the lines of the Vdovichenko paper, which is targeted at directly expressing $Z(x)$ as an infinite sum over configurations of signed loops. As we go along, we carefully address the issues of interpretation and convergence of this sum, mentioned above. In particular, Theorem 2.11, the key to the convergence issue, is to the best of our knowledge a new result.

Moreover, for our applications of the combinatorial results to the Ising model, it turns out to be necessary to allow crossing edges. This means that our function $Z(x)$ (and hence also Theorem 2.10) is not quite the same as the one considered in the literature so far. In particular, our $Z(x)$ need not be proportional to the Ising partition function for the graph $\mathcal{G}$.

2.2 Proofs of the combinatorial identities

We now turn to the proof of our main Theorems 2.9 and 2.10. Recall that here $\mathcal{G} = (V, E)$ is a general finite graph embedded in the plane, potentially containing crossing edges or additional edges. The proof proceeds in a number of steps. In the first step, detailed in Section 2.2.1, we will identify each even subgraph (V, F) of $\mathcal{G}$ with a number of edge-disjoint collections of loops, and show that the sum of their weights yields precisely the contribution of F to the generating function $Z(x)$. In the second step, in Section 2.2.2, we will explain the conditions under which we can express $Z(x)$ in terms of $\sum_r f_r(x)$ and $\det\bigl(\mathrm{I} - \Lambda(x)\bigr)$, under the assumption that the weights of all remaining configurations of loops in $\mathcal{G}$ of total length r cancel each other. The proof of this assumption, the last step, is carried out in Section 2.2.3.

2.2.1 Expansion into collections of edge-disjoint loops

We will be concerned with crossings of loops and walks, and we need to carefully establish the relevant definitions first. More specifically, we will consider collections $\{\ell_1, \dots, \ell_s\}$ of loops with the properties that all loops $\ell_1, \dots, \ell_s$ are edge-disjoint, and no two loops in the collection visit a common edge. We call these *edge-disjoint collections* of loops. Intuitively it may be clear what we mean by a crossing of such edge-disjoint loops, but some care is needed, so we will now give the precise definitions.

First, consider two walks (u, v, w) and (x, y, z) in $\mathcal{G}$, and let A be the union of the two half-lines $\{v + t(u - v) \colon t \geq 0\}$ and $\{v + t(w - v) \colon t \geq 0\}$. We say that (u, v, w) and (x, y, z) *cross each other at the vertex* v if $v = y$ and the vertices x and z do not lie in the same infinite component of the complement of A in the plane.

Now let $\ell_1 = (u_0, \dots, u_{n-1})$ and $\ell_2 = (v_0, \dots, v_{m-1})$ be two loops that form an edge-disjoint pair $\{\ell_1, \ell_2\}$. By $C_V(\ell_1, \ell_2)$ we denote the number of pairs (i, j), where $0 \leq i < n$ and $0 \leq j < m$, such that the walks (u_{i-1}, u_i, u_{i+1}) and (v_{j-1}, v_j, v_{j+1}) cross each other at u_i. We call $C_V(\ell_1, \ell_2)$ the number of *vertex crossings* between ℓ_1 and ℓ_2. Similarly, we define the number of *edge crossings* between ℓ_1 and ℓ_2, denoted $C_E(\ell_1, \ell_2)$, as the number of pairs (i, j) such that $0 \leq i < n$ and $0 \leq j < m$, and the edges $u_i u_{i+1}$ and $v_j v_{j+1}$ cross each other in $\mathcal{G}$.

We also need to formally define the number of times a loop crosses itself, so consider an edge-disjoint loop $\ell = (v_0, \ldots, v_{n-1})$. We define the number of *vertex self-crossings* of ℓ, denoted $C_V(\ell)$, as the number of pairs (i, j), where $0 \leq i < j < n$, such that (v_{i-1}, v_i, v_{i+1}) and (v_{j-1}, v_j, v_{j+1}) cross each other at the vertex v_i. The number of *edge self-crossings* of ℓ, denoted $C_E(\ell)$, is defined as the number of pairs (i, j) such that $0 \leq i < j < n$, and the edges $v_i v_{i+1}$ and $v_j v_{j+1}$ cross each other in the graph $\mathcal{G}$.

As was already mentioned in the introduction, Whitney's formula [56] says that the sign of an edge-disjoint loop is -1 if the loop crosses itself an odd number of times, and $+1$ otherwise. In other words, we have

$$\operatorname{sgn}(\ell) = (-1)^{C_V(\ell)+C_E(\ell)}$$

if ℓ is edge-disjoint. We now simply define the sign of an edge-disjoint collection of loops $\{\ell_1, \ldots, \ell_s\}$ as

$$\operatorname{sgn}\{\ell_1, \ldots, \ell_s\} = \prod_{i=1}^{s} \operatorname{sgn}(\ell_i) = (-1)^{\sum_{i=1}^{s}\{C_V(\ell_i)+C_E(\ell_i)\}}. \tag{2.15}$$

If $F \subset E$ is even, we can decompose F into an edge-disjoint collection of loops in such a way, that the union of all edges traversed by the loops is F (one way to find such a decomposition is given in the proof of Proposition 2.12 below). This decomposition is in general not unique. We write $\mathcal{D}(F)$ for the set of all possible edge-disjoint decompositions of F, and recall that $C(F)$ denotes the number of unordered pairs of edges in F that cross each other.

Proposition 2.12. *For all even subsets F of E we have that*

$$\sum_{\{\ell_1,\ldots,\ell_s\}\in\mathcal{D}(F)} \operatorname{sgn}\{\ell_1, \ldots, \ell_s\} = (-1)^{C(F)}.$$

Proof. Let $\{\ell_1, \ldots, \ell_s\}$ be an edge-disjoint collection of loops. Since $\ell_1, \ldots, \ell_s$ are all closed loops in the plane, any two distinct loops ℓ_i and ℓ_j from the collection necessarily cross each other an even number of times. That is, $C_V(\ell_i, \ell_j) + C_E(\ell_i, \ell_j)$ is even for all $i \neq j$. Therefore,

$$\operatorname{sgn}\{\ell_1, \ldots, \ell_s\} = (-1)^{\sum_{1\leq i\leq s}\{C_V(\ell_i)+C_E(\ell_i)\}+\sum_{1\leq i<j\leq s}\{C_V(\ell_i,\ell_j)+C_E(\ell_i,\ell_j)\}}.$$

Furthermore, if $\{\ell_1, \ldots, \ell_s\} \in \mathcal{D}(F)$, then clearly the total number of edge crossings occurring among the loops must coincide with $C(F)$, that is,

$$C(F) = \sum_{1\leq i\leq s} C_E(\ell_i) + \sum_{1\leq i<j\leq s} C_E(\ell_i, \ell_j).$$

Hence, it suffices to prove that

$$\sum_{\{\ell_1,\ldots,\ell_s\}\in\mathcal{D}(F)} (-1)^{\sum_{1\leq i\leq s} C_V(\ell_i)+\sum_{1\leq i<j\leq s} C_V(\ell_i,\ell_j)} = 1. \tag{2.16}$$

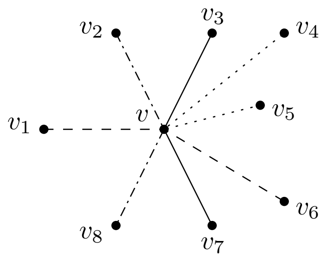

Figure 2.3: The neighbours $v_1, v_2, \dots$ of a vertex v in an even subgraph (V, F) are ordered in a clockwise fashion around v. Two neighbours that are connected by edges drawn in the same line style are paired to each other (see the text).

Let V_F denote the set of vertices in V whose degree in the subgraph (V, F) is nonzero, and let $\deg(v, F)$ denote the degree of v in (V, F). For $v \in V_F$ of degree $2k$, write $v_1, \dots, v_{2k}$ for the endpoints of the edges in F that are incident to v. Assume that these vertices are ordered in a clockwise manner around v, starting from the lexicographically smallest one (see Figure 2.3). Denote by $\mathcal{P}_v(F)$ the collection of partitions of the vertices $v_1, \dots, v_{2k}$ into sets of size 2. We call these partitions *pairings at the vertex* v. We write

$$\mathcal{P}(F) = \prod_{v \in V_F} \mathcal{P}_v(F),$$

and call an element of $\mathcal{P}(F)$ a *pairing* associated with the subgraph (V, F).

We have a natural 1–1 correspondence between $\mathcal{P}(F)$ and $\mathcal{D}(F)$. Indeed, starting from any vertex $v \in V_F$ and any $i \in \{1, \dots, \deg(v, F)\}$, the pairing $\pi \in \mathcal{P}(F)$ defines a unique closed walk $(u_0, \dots, u_{n-1})$ with the properties that $u_0 = v_i$, $u_1 = v$, and for all j, u_{j-1} is paired with u_{j+1} at the vertex u_j. Continuing this way, and replacing each closed walk obtained by the corresponding loop, yields an edge-disjoint collection $\{\ell_1, \dots, \ell_s\} \in \mathcal{D}(F)$. It is easy to see that this defines a bijective relation between $\mathcal{P}(F)$ and $\mathcal{D}(F)$.

Using this bijection, we can express the sum in (2.16) equally well as a sum over all pairings. More precisely, for $\pi \in \mathcal{P}(F)$ let π_v denote the pairing it induces at the vertex $v \in V_F$, and write $C_v(\pi_v)$ for the number of crossings at the vertex v introduced by this pairing. We call π_v *even* (*odd*) if $C_v(\pi_v)$ is even (odd). Note that (2.16) is equivalent to

$$\sum_{\pi \in \mathcal{P}(F)} (-1)^{\sum_{v \in V_F} C_v(\pi_v)} = \prod_{v \in V_F} \sum_{\pi_v \in \mathcal{P}_v(F)} (-1)^{C_v(\pi_v)} = 1,$$

from which we see that it suffices to prove that for all $v \in V_F$, the number of even pairings π_v exceeds the number of odd pairings π_v by 1.

We prove this by induction on the degree $2k$ of v. Write N_k^+ and N_k^- for the numbers of even, resp. odd, pairings for v of degree $2k$. For $k = 1$, we clearly have

$N_k^+ = 1$ and $N_k^- = 0$. Now let $k > 1$, and suppose that we pair the vertex v_1 with v_i at v. Next pair the remaining $2k-2$ neighbours v_j of v in all possible ways. For even i, there is an even number of j in between 1 and i, and therefore the pairing we obtain will be even if and only if the pairing of the remaining $2k-2$ vertices is even (see Figure 2.3). Likewise, for odd i, the obtained pairing will be even if and only if the pairing of the remaining vertices is odd. Since we have k even values for i, and $k-1$ odd values, this gives $N_k^+ = kN_{k-1}^+ + (k-1)N_{k-1}^-$ and $N_k^- = kN_{k-1}^- + (k-1)N_{k-1}^+$. Hence by the induction hypothesis, $N_k^+ - N_k^- = 1$. □

From Proposition 2.12 we will now obtain the first main result. Recall that

$$Z(x) = \sum_{\text{even } F\subset E} (-1)^{C(F)} \prod_{e\in F} x_e.$$

The case $F = \emptyset$ is treated separately: by convention it contributes 1 to the sum. Hence, Proposition 2.12 implies that

$$Z(x) = 1 + \sum_{\substack{\text{even } F\subset E:\\ F\neq\emptyset}} \sum_{\{\ell_1,\dots,\ell_s\}\in\mathcal{D}(F)} \operatorname{sgn}\{\ell_1,\dots,\ell_s\} \prod_{e\in F} x_e.$$

Recall that the multiplicity $m(\ell)$ of an edge-disjoint loop ℓ is 1. Therefore, using (2.15) and the definition (2.7) of the weight of a loop, we can write

$$Z(x) = 1 + \sum_{\substack{\text{even } F\subset E:\\ F\neq\emptyset}} \sum_{\{\ell_1,\dots,\ell_s\}\in\mathcal{D}(F)} \prod_{i=1}^{s} w(\ell_i;x).$$

Since this is a finite sum, we do not need to worry about the order of summation, so we have

$$Z(x) = 1 + \sum_{r=1}^{\infty} \sum_{\substack{\text{even } F\subset E:\\ |F\setminus E_A|=r}} \sum_{\{\ell_1,\dots,\ell_s\}\in\mathcal{D}(F)} \prod_{i=1}^{s} w(\ell_i;x).$$

If we now denote by $\mathcal{D}_r$ the set consisting of all those edge-disjoint collections of loops $\{\ell_1,\dots,\ell_s\}$ for which the total length $\sum_{i=1}^{s} r(\ell_i)$ is r, we see that we have established the following theorem:

Theorem 2.13.

$$Z(x) = 1 + \sum_{r=1}^{\infty} \sum_{\{\ell_1,\dots,\ell_s\}\in\mathcal{D}_r} \prod_{i=1}^{s} w(\ell_i;x).$$

2.2.2 Extension to all loop configurations

In Theorem 2.13, we have expressed the generating function $Z(x)$ as a sum over all edge-disjoint collections of loops in $\mathcal{G}$. In this section, we will see that if the edge weights are sufficiently small, we can drop the condition that the loops have to be edge-disjoint, and sum instead over all possible loop configurations in $\mathcal{G}$. Here, a *loop configuration* is simply an ordered sequence $(\ell_1, \dots, \ell_s)$ of loops; there is no condition that loops have to be edge-disjoint, nor that two loops in the configuration have to be distinct (i.e. it is allowed that $\ell_i = \ell_j$ for some $i \neq j$, which is why we work with ordered sequences of loops now).

Write $\mathcal{C}_r$ for the collection of all loop configurations $(\ell_1, \dots, \ell_s)$ satisfying $r(\ell_1) + \dots + r(\ell_s) = r$. Some of these loop configurations will consist of distinct loops that together form an edge-disjoint collection of loops. Let $\mathcal{C}_r^*$ denote the subset of $\mathcal{C}_r$ containing only these edge-disjoint loop configurations. Observe that if $\{\ell_1, \dots, \ell_s\}$ is an edge-disjoint collection of loops, then there are $s!$ corresponding (ordered) loop configurations $(\ell_1, \dots, \ell_s)$. Therefore, by Theorem 2.13 we already have that

$$Z(x) = 1 + \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \sum_{(\ell_1,\dots,\ell_s)\in\mathcal{C}_r^*} \frac{1}{s!} \prod_{i=1}^{s} w(\ell_i; x),$$

but we claim that here we may sum over $\mathcal{C}_r$ instead of $\mathcal{C}_r^*$:

Theorem 2.14.

$$Z(x) = 1 + \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \sum_{(\ell_1,\dots,\ell_s)\in\mathcal{C}_r} \frac{1}{s!} \prod_{i=1}^{s} w(\ell_i; x).$$

Clearly, since $\mathcal{C}_r$ is a finite set for every fixed r, this result is an immediate consequence of the following proposition:

Proposition 2.15. *For all* $r > 0$,

$$\sum_{s=1}^{\infty} \sum_{(\ell_1,\dots,\ell_s)\in\mathcal{C}_r\setminus\mathcal{C}_r^*} \frac{1}{s!} \prod_{i=1}^{s} w(\ell_i; x) = 0.$$

The proof of Proposition 2.15 is involved, and we postpone it to Section 2.2.3. For now, we assume that Proposition 2.15 and hence Theorem 2.14 hold, and explain how Theorems 2.9 and 2.10 follow from this.

Proof of Theorem 2.9. By splitting the sum over the set of loop configurations $\mathcal{C}_r$ in Theorem 2.14 according to the lengths of the individual loops, and using (2.12), we can write

$$\begin{aligned} Z(x) &= 1 + \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \frac{1}{s!} \sum_{r_1+\dots+r_s=r} \prod_{i=1}^{s} \Big(\sum_{\ell\in\mathcal{L}_{r_i}} w(\ell; x) \Big) \\ &= 1 + \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \frac{1}{s!} \sum_{r_1+\dots+r_s=r} \prod_{i=1}^{s} f_{r_i}(x). \end{aligned} \tag{2.17}$$

Now suppose that, given x, there exist $\rho \in (0,1)$ and $C < \infty$ such that

$$|f_r(x)| \leq C\rho^r \qquad \text{for all } r. \tag{2.18}$$

For future reference, we note that this condition is implied by the stronger condition that

$$\sum_{\ell\in\mathcal{L}_r}|w(\ell;x)| \le C\rho^r \qquad \text{for all } r. \tag{2.19}$$

Under condition (2.18), if we write $h(r,s)$ for the summand in (2.17), we have

$$|h(r,s)| = \left|\frac{1}{s!}\sum_{r_1+\cdots+r_s=r}\prod_{i=1}^{s} f_{r_i}(x)\right| \le \frac{C^s}{s!}\binom{r-1}{s-1}\rho^r,$$

and thus

$$\sum_{s=1}^{\infty}\sum_{r=1}^{\infty}|h(r,s)| \le \sum_{s=1}^{\infty}\frac{C^s}{s!}\sum_{r=s}^{\infty}\binom{r-1}{s-1}\rho^r = \exp\left(\frac{C\rho}{1-\rho}\right) - 1.$$

Hence, we can apply Fubini's theorem to interchange the order of summation over r and s in (2.17), which yields

$$Z(x) = 1 + \sum_{s=1}^{\infty}\frac{1}{s!}\sum_{r=1}^{\infty}\sum_{r_1+\cdots+r_s=r}\prod_{i=1}^{s} f_{r_i}(x).$$

Note that under condition (2.18), $\sum_r f_r(x)$ is absolutely convergent. We now apply Mertens' theorem, which says that if a series $\sum_r a_r$ converges absolutely, and the series $\sum_r b_r$ converges, then their Cauchy product converges to $(\sum_r a_r)(\sum_r b_r)$. In particular, by induction, the s-fold Cauchy product of the series $\sum_r a_r$ with itself, which is $\sum_r \sum_{r_1+\cdots+r_s=r} a_{r_1}a_{r_2}\cdots a_{r_s}$, converges to $(\sum_r a_r)^s$. Applying this with $a_r = f_r(x)$, we obtain

$$Z(x) = 1 + \sum_{s=1}^{\infty}\frac{1}{s!}\left(\sum_{r=1}^{\infty} f_r(x)\right)^s = \exp\left(\sum_{r=1}^{\infty} f_r(x)\right). \tag{2.20}$$

Observe that this result holds already under the weaker of the two conditions (2.18) and (2.19), but that under the stronger condition (2.19), the loop weights can in fact be summed in any order. We will now show that the condition of Theorem 2.9 implies (2.19). Indeed, under the condition of Theorem 2.9, there exists $\rho \in (0,1)$ such that $(\Delta-1)|x_e| \le \rho$ for all edges $e \in E$. Observing that if a loop takes a step, then there are at most $\Delta - 1$ possibilities for the next step, this implies that the sum of $|w(\ell;x)|$ over all loops ℓ of n steps is bounded by $|V|\rho^n$. Since a loop of length r takes at least r steps, summing over $n \ge r$ yields (2.19). □

Proof of Theorem 2.10. Recall definition (2.13) of the entries of $\Lambda(x)$, which are indexed by the directed representative edges of $\mathcal{G}$. We can interpret this matrix as a transition matrix for non-backtracking walks on the graph $\mathcal{G}'$ which is represented by $\mathcal{G}$. This represented graph $\mathcal{G}'$ can be obtained from $\mathcal{G}$ by removing every chain of additional edges from $\mathcal{G}$, and identifying the two vertices at the ends of this chain (see Figure 2.2 for an example).

Indeed, consider two directed representative edges $\overrightarrow{uv}$ and $\overrightarrow{wz} \neq \overrightarrow{vu}$ in $\mathcal{G}$, and write $\overrightarrow{uv}'$ and $\overrightarrow{wz}'$ for the corresponding directed edges in the represented graph $\mathcal{G}'$. By construction, a non-backtracking walk in $\mathcal{G}'$ can make a step from $\overrightarrow{uv}'$ to $\overrightarrow{wz}'$ if and only if v is linked to w in the graph $\mathcal{G}$, since only then will v be identified with w in $\mathcal{G}'$. This step corresponds to either a direct step from $\overrightarrow{uv}$ to $\overrightarrow{wz}$ in $\mathcal{G}$ (if $v = w$), or to a sequence of steps along the chain linking v to w. In either case, the matrix entry $\Lambda_{\overrightarrow{uv},\overrightarrow{wz}}(x)$ picks up all edge weights and turning angles associated with these steps in $\mathcal{G}$.

We can interpret this entry as describing the weight picked up by a non-backtracking walk in $\mathcal{G}'$ when it steps from $\overrightarrow{uv}'$ to $\overrightarrow{wz}'$. Viewed in this way, the entry of the matrix $\Lambda^r(x)$ indexed by $\overrightarrow{uv}$ and $\overrightarrow{wz}$ is equal to the sum of the weights of all non-backtracking walks in $\mathcal{G}'$ of r steps starting from $\overrightarrow{uv}'$ and ending on $\overrightarrow{wz}'$. In particular, the sum of the diagonal entries of $\Lambda^r(x)$ is equal to the sum of the weights of all non-backtracking walks in $\mathcal{G}'$ of r steps starting and ending on the same directed edge.

Now consider a loop ℓ of length r in $\mathcal{G}$. Note that it is possible to start traversing this loop from each step it takes along a representative edge in two directions. Mapping the walks thus obtained to the represented graph $\mathcal{G}'$ yields precisely $2r/m(\ell)$ different non-backtracking walks of r steps in $\mathcal{G}'$ that start and end on the same directed edge. By (2.7), (2.6) and (2.12), it now follows that

$$\operatorname{tr} \Lambda^r(x) = -2r f_r(x),$$

where the minus sign comes from the minus sign in the definition (2.6) of the sign of a loop in terms of its winding angle. Expressed in the eigenvalues $\lambda_i(x)$ of $\Lambda(x)$, we therefore have that

$$f_r(x) = -\frac{1}{2r} \sum_i \lambda_i^r(x). \tag{2.21}$$

In particular, condition (2.18) is satisfied if $\rho(x) = \max_i |\lambda_i(x)| < 1$, so in this case the same argument as in the proof of Theorem 2.9 yields (2.20). Moreover, if $\rho(x) < 1$, then using (2.21) we can write

$$Z(x) = \exp\biggl(-\frac{1}{2} \sum_{r=1}^{\infty} \sum_i \frac{\lambda_i^r(x)}{r}\biggr) = \exp\biggl(-\frac{1}{2} \sum_i \sum_{r=1}^{\infty} \frac{\lambda_i^r(x)}{r}\biggr),$$

and since $\sum_{r=1}^{\infty} u^r/r = -\ln(1-u)$ if $|u| < 1$, we conclude that

$$Z(x) = \prod_i \bigl(1 - \lambda_i(x)\bigr)^{1/2} = \det{}^{\frac{1}{2}} \bigl(\mathrm{I} - \Lambda(x)\bigr). \qquad \square$$

2.2.3 Cancellation of non-edge-disjoint loop configurations

We now turn to the missing step in the proofs of Theorems 2.9 and 2.10, which is the proof of Proposition 2.15. That is, we must show that the weights of all loop configurations $(\ell_1, \dots, \ell_s)$ which are not edge-disjoint and satisfy $r(\ell_1) + \dots + r(\ell_s) = r$ for

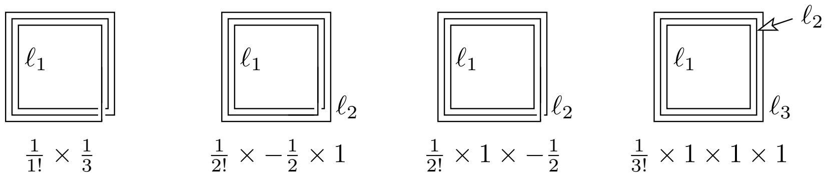
ℓ_1
$\frac{1}{1!} \times \frac{1}{3}$
ℓ_1
ℓ_2
$\frac{1}{2!} \times -\frac{1}{2} \times 1$
ℓ_1
ℓ_2
$\frac{1}{2!} \times 1 \times -\frac{1}{2}$
ℓ_2
ℓ_1
ℓ_3
$\frac{1}{3!} \times 1 \times 1 \times 1$

Figure 2.4: Four loop configurations on the same vertices and edges, where the traversals of the same edge have been drawn slightly apart to make them discernible. The factors $\frac{1}{s!}\prod_{i=1}^{s} \operatorname{sgn}(\ell_i)/m(\ell_i)$ are spelled out below each loop configuration to show that the sum of their signed weights is 0.

a given r, cancel each other. What complicates matters here is the fact that these loop configurations do not cancel each other one by one, see for example Figure 2.4.[1] Our strategy of the proof is to map loop configurations to so-called *labelled* loop configurations, which do cancel each other one by one, and show that this implies cancellation of the unlabelled loop configurations for combinatorial reasons.

We will therefore start by introducing the notion of a *labelled* loop, and work our way from there towards the notion of a labelled loop configuration, and the proof of their cancellation. In words, a labelled loop is a loop with a label attached to each step it takes, where the labels are distinct positive integers. For periodic loops, the first step is repeated after completing a period, and we require that the label of the first step of the loop is smaller than the label associated with each of these repetitions.

Formally, a labelled loop $\ell^\diamond$ is a sequence $(v_0, a_0, v_1, a_1, \ldots, v_{n-1}, a_{n-1})$ satisfying the following conditions:

L1 $\ell = (v_0, \ldots, v_{n-1})$ is a loop;

L2 $(a_0, a_1, \ldots, a_{n-1})$ is a sequence of distinct positive integers, called the *labelling* of the loop;

L3 if ℓ is periodic, i.e. $m(\ell) > 1$, then a_0 is smaller than $a_{kn/m(\ell)}$ for all $k \in \{1, 2, \ldots, m(\ell) - 1\}$.

We call the number a_i the *label on step* $i + 1$ of the loop ℓ; we also regard it as a label assigned to the edge $v_i v_{i+1}$. We will use the superscript $\diamond$ for labelled loops, and the unlabelled loop corresponding to a labelled loop will consistently be denoted by dropping this superscript: if $\ell^\diamond$ is a labelled loop, then ℓ is the corresponding unlabelled loop, and so on.

Observe that one of the effects of labelling loops is that it breaks the periodicity of periodic loops: sequences representing labelled loops cannot be periodic. Therefore, if ℓ is periodic, we do not assign to the labelled loop $\ell^\diamondsuit = (v_0, a_0, \dots, v_{n-1}, a_{n-1})$ the same weight as to its unlabelled counterpart. Instead, we define the weights of labelled loops in general by

---

[1] A picture of the same configurations appears in [20] to point out the error in Vdovichenko's paper; it is crucial here to take the multiplicities of the loops into account.

$$w(\ell^\diamondsuit; x) = \operatorname{sgn}(\ell) \prod_{i=0}^{n-1} x_{v_i v_{i+1}}, \tag{2.22}$$

where the sign is defined in terms of the winding angle of ℓ by (2.6), as before. Note that this weight is actually independent of the particular labelling of the loop, and that $w(\ell^\diamondsuit; x) = m(\ell) w(\ell; x)$.

We write $n(\ell)$ for the number of steps of a loop ℓ (recall that this is not necessarily the same as the *length* $r(\ell)$ of the loop). By a *labelled loop configuration* we mean a collection $\{\ell_1^\diamondsuit, \dots, \ell_s^\diamondsuit\}$ of labelled loops, in which all labels are distinct and take values from the set $\bigl\{1, 2, \dots, \sum_{i=1}^s n(\ell_i)\bigr\}$. In particular, any loop configuration $(\ell_1, \dots, \ell_s)$ can be turned into a labelled loop configuration by attaching a label to every step of every loop in such a way that condition **L3** above is fulfilled for every labelled loop obtained, and all labels $1, 2, \dots, \sum_{i=1}^s n(\ell_i)$ are used.

Now fix r and n, and consider a loop configuration $(\ell_1, \dots, \ell_s)$ which is not edge-disjoint and satisfies $\sum_{i=1}^s r(\ell_i) = r$ and $\sum_{i=1}^s n(\ell_i) = n$. Let t denote the number of distinct loops in $(\ell_1, \dots, \ell_s)$, and write $k_1, \dots, k_t$ for the respective number of times each of them occurs, so that $k_1 + \dots + k_t = s$. Consider the collection of all labelled loop configurations $\{\ell_1^\diamondsuit, \dots, \ell_s^\diamondsuit\}$ that can be obtained from $(\ell_1, \dots, \ell_s)$ by labelling the loops, as described above. For a periodic loop ℓ_i, only one of the rotations of its labelling, rotated over a multiple of the smallest period, satisfies condition **L3**. Furthermore, interchanging the labellings of two identical loops ℓ_i and ℓ_j ($\ell_i = \ell_j$ but $i \neq j$) yields the same labelled loop configuration. Therefore, the number of labelled loop configurations we obtain from $(\ell_1, \dots, \ell_s)$ is precisely

$$\frac{n!}{\prod_{i=1}^s m(\ell_i) \prod_{i=1}^t k_i!}.$$

We assign to each of these labelled loop configurations the same weight equal to $\prod_{i=1}^s w(\ell_i^\diamondsuit; x)$, where we use the fact that according to the definition (2.22), $w(\ell_i^\diamondsuit; x)$ does not depend on the actual labelling. Then the total weight of all labelled loop configurations associated with $(\ell_1, \dots, \ell_s)$ is

$$\frac{n!}{\prod_{i=1}^s m(\ell_i) \prod_{i=1}^t k_i!} \prod_{i=1}^s w(\ell_i^\diamondsuit; x) = \frac{n!}{\prod_{i=1}^t k_i!} \prod_{i=1}^s w(\ell_i; x).$$

We claim that this is exactly $n!$ times the total weight that all the permutations of the loop configuration $(\ell_1, \dots, \ell_s)$ contribute to the sum in Proposition 2.15. Indeed, there are precisely

$$\frac{s!}{\prod_{i=1}^s k_i!}$$

such permutations, and the weight each of them contributes to the sum is

$$\frac{1}{s!}\prod_{i=1}^{s} w(\ell_i; x).$$

We conclude that to prove Proposition 2.15, it suffices to show that for given n and r, the weights of all labelled loop configurations $\{\ell_1^\diamondsuit, \ldots, \ell_s^\diamondsuit\}$ such that $\sum_{i=1}^{s} n(\ell_i) = n$, $\sum_{i=1}^{s} r(\ell_i) = r$ and $(\ell_1, \ldots, \ell_s)$ is not edge-disjoint, sum to 0. Write $\mathcal{C}_{n,r}^\diamondsuit$ for this collection of labelled loop configurations. We will now prove the desired cancellation of weights, and hence Proposition 2.15, by finding a bijection $g\colon \mathcal{C}_{n,r}^\diamondsuit \to \mathcal{C}_{n,r}^\diamondsuit$ which maps each labelled loop configuration to a labelled loop configuration which has a weight of the opposite sign, but with the same absolute value.

Proof of Proposition 2.15. Before we go into the formal details of the bijection, let us give an informal description of how it will work. Consider a labelled loop configuration $\{\ell_1^\diamondsuit, \ldots, \ell_s^\diamondsuit\} \in \mathcal{C}_{n,r}^\diamondsuit$, and let $E^\diamondsuit$ be the set of edges in $\mathcal{G}$ that are assigned more than one label in this configuration. Find the smallest of all the labels that are assigned to the edges in $E^\diamondsuit$, let a be this label, and let uv be the edge to which this label is assigned. Next, find the second smallest label b which is assigned to the edge uv.

The label a labels a step of one of the loops ℓ_i. The label b either labels another step of the same loop ℓ_i, or it labels a step of a second loop ℓ_j, $i \neq j$. The bijection involves interchanging the "connections" on one side of the two steps marked a and b (either at the vertex u or at the vertex v), as illustrated in Figure 2.5. It is clear that this operation does not change the absolute value of the weight of the configuration, since the total number of steps that go through a given edge does not change. But Figure 2.5 also suggests that the operation corresponds to increasing or decreasing the number of "crossings" in the configuration by 1, which should indeed lead to a change in sign.

However, signs were formally defined in terms of winding angles, not numbers of crossings, since it is more difficult to make sense of the latter when loops are not edge-disjoint. Furthermore, we must still formally define the mapping g. We will now deal with these technical issues.

For the formal treatment of the bijection, we need to introduce some additional notation. Given a sequence $a = (a_0, \ldots, a_n)$ of arbitrary elements, we write a^{-1} for its *reversion* $a^{-1} = (a_n, a_{n-1}, \ldots, a_0)$. If $b = (b_0, \ldots, b_m)$ is another sequence of arbitrary elements, we write $a \oplus b$ for the concatenation of a with b, that is,

$$a \oplus b = (a_0, \ldots, a_n, b_0, \ldots, b_m).$$

The weight of a labelled loop configuration $\{\ell_1^\diamondsuit, \ldots, \ell_s^\diamondsuit\}$ is defined as the product of the signs of the loops $\ell_1, \ldots, \ell_s$, times the product of all the edge weights picked up by all the loops. As was anticipated above, the product of edge weights will not change under the bijection, so we will only be concerned with the product of the signs of the loops. We recall from (2.5) and (2.6) that the sign of a loop $\ell = (v_0, \ldots, v_{n-1})$

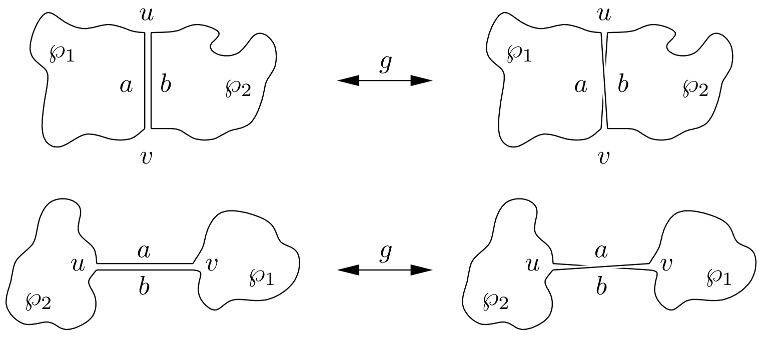

Figure 2.5: All cases that occur in the cancellation of labelled loop diagrams, as explained in the text. The curves ω_1 and ω_2 represent arbitrary walks connected to the vertices u and v.

is defined in terms of its winding angle as

$$\operatorname{sgn}(\ell) = -\exp\Big(\frac{i}{2}\alpha(\ell)\Big), \tag{2.23}$$

where the winding angle $\alpha(\ell)$ is given by

$$\alpha(\ell) = \sum_{i=0}^{n-1} \angle(v_{i+1} - v_i, v_{i+2} - v_{i+1}). \tag{2.24}$$

We now define the winding angle and sign of a closed walk $(v_0, \dots, v_{n-1})$ by the exact same formulas. In particular, all rotations (i.e. circular shifts) of a loop ℓ have the same winding angle and sign. On the other hand, the reversion ℓ^{-1} of ℓ and all its rotations are traversed in the opposite direction, and therefore they all have winding angle $\alpha(\ell^{-1}) = -\alpha(\ell)$. However, since the winding angle of a loop is a multiple of 2π, we do have that

$$\operatorname{sgn}(\ell^{-1}) = \operatorname{sgn}(\ell) \qquad \text{for all closed walks } \ell. \tag{2.25}$$

We call all the rotations of a loop ℓ, and all rotations of its reversion ℓ^{-1}, alternative *representations* of ℓ. All these representations have the same sign. Likewise, the rotations of a labelled loop $\ell^{\diamondsuit}$ and its reversion $(\ell^{\diamondsuit})^{-1}$ will be called *representations* of this labelled loop.

We also need to define the winding angle for walks in $\mathcal{G}$ which are not loops. Note that a walk $\omega = (v_0, \dots, v_{n-1})$ is not a loop if $v_0 v_{n-1} \notin E$, $v_0 = v_{n-2}$, or $v_1 = v_{n-1}$. If we follow such a walk from v_0 to v_{n-1}, we turn through $n-2$ angles, and it is natural to define the winding angle of ω by

$$\alpha(\omega) = \sum_{i=0}^{n-3} \angle(v_{i+1} - v_i, v_{i+2} - v_{i+1}).$$

We now have all the notation we need to define and analyse the bijection formally. So consider a labelled loop configuration $\{\ell_1^\diamondsuit, \dots, \ell_s^\diamondsuit\} \in \mathcal{C}_{n,r}^\diamondsuit$, and define $E^\diamondsuit$, a, b and uv as above. We will now explain to which labelled loop configuration our configuration $\{\ell_1^\diamondsuit, \dots, \ell_s^\diamondsuit\}$ is mapped by the bijection, and prove that the image has the opposite sign, and hence the opposite weight. There are three possible cases to consider, which are illustrated in Figure 2.5.

Case 1: *The labels a and b belong to different labelled loops.* Let $\ell_i^\diamondsuit$ be the labelled loop containing label a, and let $\ell_j^\diamondsuit$ be the labelled loop containing label b. Then these labelled loops have representations of the form $\hat\ell_i^\diamondsuit = (u,a,v) \oplus \omega_1^\diamondsuit$ and $\hat\ell_j^\diamondsuit = (u,b,v) \oplus \omega_2^\diamondsuit$, respectively, where $\omega_1^\diamondsuit$ and $\omega_2^\diamondsuit$ are walks interspersed with labels. We can now form the combined representation

$$\hat\ell_{ij}^\diamondsuit = (u,a,v) \oplus \omega_1^\diamondsuit \oplus (u,b,v) \oplus \omega_2^\diamondsuit$$

of a new labelled loop $\ell_{ij}^\diamondsuit$. Our bijection maps $\{\ell_1^\diamondsuit, \dots, \ell_s^\diamondsuit\}$ to the labelled loop configuration

$$\{\ell_1^\diamondsuit, \dots, \ell_s^\diamondsuit, \ell_{ij}^\diamondsuit\} \setminus \{\ell_i^\diamondsuit, \ell_j^\diamondsuit\}.$$

To see that this labelled loop configuration has the opposite sign compared to the sign of its pre-image $\{\ell_1^\diamondsuit, \dots, \ell_s^\diamondsuit\}$, note that by (2.23)–(2.25),

$$\begin{aligned} \operatorname{sgn}(\ell_i)\operatorname{sgn}(\ell_j) &= \operatorname{sgn}(\hat\ell_i)\operatorname{sgn}(\hat\ell_j) = \exp\Bigl(\frac{i}{2}\alpha(\hat\ell_i) + \frac{i}{2}\alpha(\hat\ell_j)\Bigr) \\ &= \exp\Bigl(\frac{i}{2}\alpha(\hat\ell_{ij})\Bigr) = -\operatorname{sgn}(\hat\ell_{ij}) = -\operatorname{sgn}(\ell_{ij}). \end{aligned}$$

Case 2: *The labels a and b are on steps of the same labelled loop taken in the same direction.* This case is the reverse of Case 1. The labels a and b are in a labelled loop $\ell_i^\diamondsuit$ which has a representation of the form

$$\hat\ell_i^\diamondsuit = (u,a,v) \oplus \omega_1^\diamondsuit \oplus (u,b,v) \oplus \omega_2^\diamondsuit.$$

From this we obtain the representations $(u,a,v) \oplus \omega_1^\diamondsuit$ and $(u,b,v) \oplus \omega_2^\diamondsuit$ of two new labelled loops $\ell_{i1}^\diamondsuit$ and $\ell_{i2}^\diamondsuit$. The bijection maps $\{\ell_1^\diamondsuit, \dots, \ell_s^\diamondsuit\}$ to the labelled loop configuration

$$\{\ell_1^\diamondsuit, \dots, \ell_s^\diamondsuit, \ell_{i1}^\diamondsuit, \ell_{i2}^\diamondsuit\} \setminus \{\ell_i^\diamondsuit\}.$$

The same argument as in Case 1 shows that $\operatorname{sgn}(\ell_{i1})\operatorname{sgn}(\ell_{i2}) = -\operatorname{sgn}(\ell_i)$.

Case 3: *The labels a and b are on steps of the same labelled loop taken in opposite directions.* In this case the labels a and b are in a labelled loop $\ell_i^\diamondsuit$ which has a representation of the form

$$\hat\ell_i^\diamondsuit = (u,a,v) \oplus \omega_1^\diamondsuit \oplus (v,b,u) \oplus \omega_2^\diamondsuit.$$

From this we can construct the representation

$$\hat\ell^\diamondsuit = (u,a,v) \oplus (\omega_1^\diamondsuit)^{-1} \oplus (v,b,u) \oplus \omega_2^\diamondsuit$$

of a new labelled loop $\ell^\diamondsuit$. The bijection maps $\{\ell_1^\diamondsuit, \dots, \ell_s^\diamondsuit\}$ to the labelled loop configuration

$$\{\ell_1^\diamondsuit, \dots, \ell_s^\diamondsuit, \ell^\diamondsuit\} \setminus \{\ell_i^\diamondsuit\}.$$

To verify that these loop configurations have opposite signs, observe that

$$\alpha(\hat{\ell}_i) = \alpha\big((u,v) \oplus \omega_1 \oplus (v,u)\big) + \alpha\big((v,u) \oplus \omega_2 \oplus (u,v)\big), \tag{2.26}$$

and likewise

$$\alpha(\hat{\ell}) = \alpha\big((u,v) \oplus \omega_1^{-1} \oplus (v,u)\big) + \alpha\big((v,u) \oplus \omega_2 \oplus (u,v)\big), \tag{2.27}$$

where ω_1 and ω_2 are the walks obtained from $\omega_1^\diamondsuit$ and $\omega_2^\diamondsuit$ by dropping the labels. Now notice that upon reversion,

$$\alpha\big((u,v) \oplus \omega_1 \oplus (v,u)\big) = -\alpha\big((u,v) \oplus \omega_1^{-1} \oplus (v,u)\big). \tag{2.28}$$

Furthermore, it is not difficult to see that

$$\alpha\big((u,v) \oplus \omega_1 \oplus (v,u)\big) = 2m\pi + \pi \qquad \text{for some } m \in \mathbb{Z}.$$

Together with (2.26), (2.27) and (2.28), this implies

$$\frac{\operatorname{sgn}(\ell_i)}{\operatorname{sgn}(\ell)} = \frac{\operatorname{sgn}(\hat{\ell}_i)}{\operatorname{sgn}(\hat{\ell})} = \exp\Big(\frac{i}{2}\alpha(\hat{\ell}_i) - \frac{i}{2}\alpha(\hat{\ell})\Big) = -1.$$

We conclude that in all three cases, the labelled loop configuration $\{\ell_1^\diamondsuit, \dots, \ell_s^\diamondsuit\}$ is mapped to a labelled loop configuration of opposite weight. From the explicit descriptions given above, it is not difficult to see that the mapping is bijective. As we have explained above, this implies Proposition 2.15. ☐

2.3 Bound on the operator norm of the transition matrix

In this section, we will prove Theorem 2.11. To this end, let $\mathcal{G} = (V, E)$ be a fixed finite rectangle in $\mathbb{Z}^2$ with no additional edges (i.e. $E_A = \emptyset$). Without loss of generality, we may assume that the vertex set is

$$V = \{0, 1, \dots, M-1\} \times \{0, 1, \dots, N-1\}. \tag{2.29}$$

Since we are on the square lattice, directed edges can point in only 4 directions, and we now introduce some convenient notation for this specific case. We write $v{\uparrow}$, $v{\downarrow}$, $v{\rightarrow}$ and $v{\leftarrow}$ for the directed edges from v to, respectively, $v + (0,1)$, $v - (0,1)$, $v + (1,0)$ and $v - (1,0)$. We also write ${\uparrow}v$ for the directed edge pointing from $v - (0,1)$ to v, and define ${\downarrow}v$, ${\rightarrow}v$, ${\leftarrow}v$ analogously.

Given a vector of edge weights $x = (x_e)_{e\in E}$, $\Lambda(x)$ is the transition matrix indexed by the directed edges of $\mathcal{G}$, defined by (2.13). For a vertex v not on the boundary of $\mathcal{G}$, the row of $\Lambda(x)$ indexed by ${\to}v$, for instance, has exactly 3 nonzero entries, corresponding to the 3 possible steps that a loop can take from ${\to}v$. To be precise, with $u = v - (1,0)$, these 3 entries are

$$\Lambda_{{\to}v,v{\to}}(x) = x_{uv}, \quad \Lambda_{{\to}v,v\uparrow}(x) = x_{uv}e^{i\pi/4}, \quad \Lambda_{{\to}v,v\downarrow}(x) = x_{uv}e^{-i\pi/4}.$$

Observe that most rows of $\Lambda(x)$ have exactly 3 nonzero entries. The only exceptions are the rows indexed by directed edges pointing to a vertex in $\partial\mathcal{G}$. These exceptional rows make it impossible to compute the eigenvalues of $\Lambda(x)$ directly. We will therefore make the graph periodic by connecting opposite sides, as described in Section 2.1.3, so that all vertices can be treated alike, and then bound the eigenvalues of $\Lambda(x)$ in terms of those of the periodic graph (or equivalently, a graph wrapped on a torus).

To be precise, we first extend our graph $\mathcal{G}$ to a graph $\mathcal{G}^{\circledcirc}$, by adding edges and vertices as shown in Figure 2.2 (left). Note that this adds directed *representative* edges $v{\to}$ and ${\leftarrow}v$ for every vertex v on the right boundary of $\mathcal{G}$, and $v{\uparrow}$ and ${\downarrow}v$ for v on the top boundary. All other edges that are added are considered as *additional* edges in the graph $\mathcal{G}^{\circledcirc}$. Henceforth, when we work on the graph $\mathcal{G}^{\circledcirc}$, computations will be performed modulo M and N in the two respective lattice directions.

We define $\Lambda^{\circledcirc}$ as the transition matrix for the graph $\mathcal{G}^{\circledcirc}$, with specific edge weights chosen as follows: all representative edges of $\mathcal{G}^{\circledcirc}$ have edge weight 1; for the additional edges, we choose the edge weights in such a way, that the product of the edge weights along every chain of additional edges linking opposite sides of the rectangle to each other is -1. Note that by this choice, the factor -1 will exactly compensate the sign picked up by a walk which follows the chain, because of the 4 quarter-turns it makes.

Proof of Theorem 2.11. We first prove that the operator norm of the matrix $\Lambda^{\circledcirc}$ equals $\sqrt{2}+1$. To this end, assume that the rows of $\Lambda^{\circledcirc}$ are arranged in such a way, that for every vertex $v \in V$, the 4 rows indexed by ${\to}v$, ${\uparrow}v$, ${\leftarrow}v$ and ${\downarrow}v$ immediately succeed each other in this order. Let Π be the permutation matrix which permutes the columns of $\Lambda^{\circledcirc}$ so that column vd maps to column dv, for all $v \in V$ and $d \in \{\uparrow, \downarrow, \rightarrow, \leftarrow\}$.

By construction, the matrix $\Lambda^{\circledcirc}\Pi$ with the permuted columns is a block-diagonal matrix, since the 4 rows indexed by the directed edges pointing to v are matched along the diagonal with the 4 columns indexed by the directed edges pointing out from v. By considering the turning angles, it is easy to see that each 4×4 block is equal to the Hermitian matrix

$$A = \begin{bmatrix} 1 & \exp(i\pi/4) & 0 & \exp(-i\pi/4) \\ \exp(-i\pi/4) & 1 & \exp(i\pi/4) & 0 \\ 0 & \exp(-i\pi/4) & 1 & \exp(i\pi/4) \\ \exp(i\pi/4) & 0 & \exp(-i\pi/4) & 1 \end{bmatrix},$$

which has eigenvalues $\sqrt{2}+1$ and $\sqrt{2}-1$, both of multiplicity 2.

Since A is Hermitian, its spectral radius is equal to its operator norm $\|A\|$. It follows that the operator norm of $\Lambda^{\circledcirc}\Pi$ is given by $\|A\| = \sqrt{2}+1$, and since permuting columns does not change the operator norm of a matrix, we conclude that $\|\Lambda^{\circledcirc}\| = \sqrt{2}+1$.

We will now use this fact, together with the sub-multiplicativity of the operator norm, to bound $\|\Lambda(x)\|$. To this end, let $D(x)$ be the diagonal matrix of the same dimensions as $\Lambda^{\circledcirc}$, defined as follows. For vertices v on the right boundary of $\mathcal{G}$, the diagonal entries of $D(x)$ on the rows $v\!\rightarrow$ and $\leftarrow\! v$ are 0, and so are the diagonal entries on the rows $v\!\uparrow$ and $\downarrow\! v$ for v on the top boundary of $\mathcal{G}$. For all other directed edges $\overrightarrow{uv}$ in the graph $\mathcal{G}^{\circledcirc}$, the diagonal entry of $D(x)$ on row $\overrightarrow{uv}$ is equal to the edge weight x_{uv}.

Now consider the matrix $D(x)\Lambda^{\circledcirc}D(1)$, where 1 denotes the edge weight vector on $\mathcal{G}^{\circledcirc}$ with constant weight 1 on every edge. The multiplication by $D(x)$ multiplies all rows of $\Lambda^{\circledcirc}$ corresponding to directed edges $\overrightarrow{uv}$ in the graph $\mathcal{G}$ by x_{uv}, and zeroes out all rows corresponding to directed edges which are in the graph $\mathcal{G}^{\circledcirc}$, but not in the graph $\mathcal{G}$. The multiplication by $D(1)$ then zeroes out all columns of $\Lambda^{\circledcirc}$ corresponding to directed edges which are in $\mathcal{G}^{\circledcirc}$ but not in $\mathcal{G}$. In other words, $D(x)\Lambda^{\circledcirc}D(1)$ is just the matrix $\Lambda(x)$ with rows and columns of zeros added to it for every directed edge which is in $\mathcal{G}^{\circledcirc}$ but not in $\mathcal{G}$. Therefore,

$$\|\Lambda(x)\| = \|D(x)\Lambda^{\circledcirc}D(1)\| \leq \|D(x)\| \cdot \|\Lambda^{\circledcirc}\| \cdot \|D(1)\| = (\sqrt{2}+1)\|x\|_\infty.$$

The desired bound on $|f_r(x)|$ now follows from (2.21) and the facts that $\rho(x) \leq \|\Lambda(x)\|$ and the number of directed edges in $\mathcal{G}$ is bounded by $4|V|$. ☐

2.4 Proofs of the results for the Ising model

In the next sections, we will apply Theorem 2.10 to the Ising model on the square lattice $\mathbb{Z}^2$. This will lead to explicit expressions for the free energy density and two-point functions in terms of sums over loops in $\mathbb{Z}^2$ or its dual $\mathbb{Z}^{2*}$, valid all the way up to the critical point. We start with a brief review of the low- and high-temperature expansions in Section 2.4.1. Then, we will study the free energy density in Section 2.4.2, and finally the two-point functions at low and high temperatures in Sections 2.4.3 and 2.4.4, respectively.

2.4.1 Low- and high-temperature expansions of the partition function

The partition function of the Ising model is closely related to the graph generating function $Z(x)$ from Section 2.1.4. This can be seen from the *low-* and *high-temperature* expansions considered in this section. More details on these expansions and the related duality of the Ising model can be found in [52, Section II.7].

Let $\mathcal{G} = (V, E)$ be a finite rectangle in $\mathbb{Z}^2$. By $\mathcal{G}^* = (V^*, E^*)$ we shall denote the *weak dual* graph of $\mathcal{G}$, i.e. the rectangle in $\mathbb{Z}^{2*}$ whose vertices are the centres of the

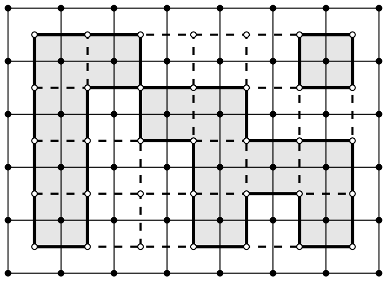

Figure 2.6: The graph $\mathcal{G}$ and its weak dual $\mathcal{G}^*$, with an even subgraph of $\mathcal{G}^*$ marked by bold edges. The spins in the gray squares have value -1, the rest have value $+1$.

bounded faces of $\mathcal{G}$ (see Figure 2.6, left). For our purposes, the low-temperature expansion is best considered in the case of positive boundary conditions. It is not difficult to see that in this case, there is a 1–1 correspondence between the even subgraphs of the weak dual $\mathcal{G}^*$ and the spin configurations in Ω^+: given $\sigma \in \Omega^+$, one obtains the corresponding even subset $F(\sigma)$ of E^* by including the edge dual to uv in $F(\sigma)$ if and only if $\sigma_u \neq \sigma_v$, for every $uv \in E$. See Figure 2.6 for an illustration.

Note that by this correspondence, if $F \subset E^*$ is even, then every edge in F separates two spins that have opposite sign. This means that adding an edge uv to F decreases $\sigma_u\sigma_v$ from $+1$ to -1, and hence has a "cost" $\exp(-2\beta)$ in the probability distribution (2.1). It follows that we can write

$$\mathbf{P}^+_{\mathcal{G},\beta}(\sigma) = \frac{\exp(\beta|E|)}{\mathcal{Z}^+_{\mathcal{G},\beta}} \prod_{e\in F(\sigma)} x_e, \qquad \sigma \in \Omega^+_{\mathcal{G}}, \tag{2.30}$$

where $x_e = \exp(-2\beta)$ for every $e \in E^*$, and

$$\mathcal{Z}^+_{\mathcal{G},\beta} = \exp(\beta|E|) \sum_{\text{even } F\subset E^*} \prod_{e\in F} x_e. \tag{2.31}$$

This is the low-temperature expansion of the partition function for positive boundary conditions. Observe that up to the factor $\exp(\beta|E|)$, this expansion takes exactly the form (2.10) of the graph generating function $Z(x)$ for the weak dual graph $\mathcal{G}^*$ (in which no edges cross each other), if we set the edge weights of all dual edges $e \in E^*$ equal to $x_e = \exp(-2\beta)$.

We now turn to the high-temperature expansion, for which we impose free boundary conditions. The expansion will be over even subgraphs of the graph $\mathcal{G}$, rather than of the weak dual $\mathcal{G}^*$. Unlike in the low-temperature expansion, these subgraphs do not have a clear geometric interpretation in terms of the spin configurations.

The high-temperature expansion starts from equation (2.2) and the observation that $\sigma_u\sigma_v$ can only take the values -1 or $+1$. Since $\exp(\pm\beta) = \cosh\beta \pm \sinh\beta$, we see that

$$\mathcal{Z}^{\text{free}}_{\mathcal{G},\beta} = (\cosh\beta)^{|E|} \sum_{\sigma\in\Omega^{\text{free}}_{\mathcal{G}}} \prod_{uv\in E} \big(1+\sigma_u\sigma_v\tanh\beta\big).$$

The next step is to expand the product over $uv \in E$. Each term in the expansion will be a product of factors obtained by choosing for each edge uv whether 1 is taken as a factor, or $\sigma_u\sigma_v\tanh\beta$, so that the expansion becomes a sum over all choices of factors for each edge uv. We can represent each choice graphically by removing the edge uv if we choose the factor 1 for this edge, and keeping uv if we choose the factor $\sigma_u\sigma_v\tanh\beta$. This gives a 1–1 correspondence between all terms in the expansion, and all $F\subset E$ (not just the even ones). Using this correspondence, and then interchanging the order of summation over σ and F, we may now write the partition function as

$$\mathcal{Z}^{\text{free}}_{\mathcal{G},\beta} = (\cosh\beta)^{|E|} \sum_{F\subset E} \sum_{\sigma\in\Omega^{\text{free}}_{\mathcal{G}}} \prod_{u\in V} \sigma_u^{\deg(u,F)} \prod_{uv\in F} x_{uv}, \tag{2.32}$$

where $x_{uv} = \tanh\beta$ for all $uv\in E$ and $\deg(u,F)$ denotes the degree of u in the graph (V,F). Note that the sum over σ vanishes unless $\deg(u,F)$ is even for all $u\in V$, in which case the sum yields simply $2^{|V|}$. Therefore,

$$\mathcal{Z}^{\text{free}}_{\mathcal{G},\beta} = 2^{|V|}(\cosh\beta)^{|E|} \sum_{\text{even } F\subset E} \prod_{uv\in F} x_{uv}. \tag{2.33}$$

Again, up to a multiplicative constant, the expansion takes exactly the form (2.10) of the graph generating function $Z(x)$, this time for the graph $\mathcal{G}$, if we set the edge weights equal to $x_{uv} = \tanh\beta$.

For future reference, note that the above expansions did not rely on the fact that the graphs are rectangles in the plane. Indeed, the low-temperature expansion can be performed for all planar graphs and the high-temperature expansion works for arbitrary graphs.

2.4.2 Free energy density

We are now going to use the bound obtained in Theorem 2.11 to prove Theorem 2.3 and it's Corollary 2.4.

Proof of Theorem 2.3. We start with the high-temperature case, so fix $\beta\in(0,\beta_c)$ and set $x=\tanh\beta$. Let $\mathcal{G}$ be a rectangle in $\mathbb{Z}^2$, and take the set of additional edges to be empty. Note that by (2.4), $x\in(0,\sqrt{2}-1)$. By (2.33),

$$\ln \mathcal{Z}^{\text{free}}_{\mathcal{G},\beta} = |V|\ln 2 + |E|\ln(\cosh\beta) + \ln Z_{\mathcal{G}}(x), \tag{2.34}$$

where $Z_{\mathcal{G}}(x)$ is the generating function for the graph $\mathcal{G}$ with edge weights equal to x. By Theorems 2.10 and 2.11, $\ln Z_{\mathcal{G}}(x)$ equals $\sum_r f_{\mathcal{G},r}(x)$, where $f_{\mathcal{G},r}(x)$ denotes the sum of the weights of all loops of length r in $\mathcal{G}$.

Consider these loops of length r in $\mathcal{G}$. For each vertex $v \in V$, let $\mathcal{L}_r^v(\mathcal{G})$ denote the collection of loops in $\mathcal{G}$ of length r for which v is the smallest vertex in the lexicographic order on $\mathbb{Z}^2$ that the loop visits. Observe that if v is at distance at least r to the boundary of $\mathcal{G}$, then $\mathcal{L}_r^v(\mathcal{G})$ can be mapped bijectively to $\mathcal{L}_r^\circ(\mathbb{Z}^2)$ by a translation on $\mathbb{Z}^2$, hence $\sum_{\ell\in\mathcal{L}_r^v(\mathcal{G})} w(\ell;x) = f_r^\circ(x)$. There are at most $|\partial\mathcal{G}|r$ vertices at a distance less than r from $\partial\mathcal{G}$, and since $|x| < 1$, for such a vertex v we have $\sum_{\ell\in\mathcal{L}_r^v(\mathcal{G})}|w(\ell;x)| \le 3^r$ (by counting non-backtracking walks). From these observations and the fact that $\lim_{\mathcal{G}\to\mathbb{Z}^2}|\partial\mathcal{G}|/|V| = 0$, it follows that

$$\lim_{\mathcal{G}\to\mathbb{Z}^2}\frac{1}{|V|}f_{\mathcal{G},r}(x) = \lim_{\mathcal{G}\to\mathbb{Z}^2}\frac{1}{|V|}\sum_{v\in V}\sum_{\ell\in\mathcal{L}_r^v(\mathcal{G})} w(\ell;x) = f_r^\circ(x) \quad \text{for all } r \ge 1.$$

Furthermore, Theorem 2.11 says that $|f_{\mathcal{G},r}(x)| \le 2|V|r^{-1}(\sqrt{2}+1)^r x^r$. Therefore, by dominated convergence and Theorem 2.10,

$$\lim_{\mathcal{G}\to\mathbb{Z}^2}\frac{1}{|V|}\ln Z_{\mathcal{G}}(x) = \lim_{\mathcal{G}\to\mathbb{Z}^2}\sum_{r=1}^{\infty}\frac{1}{|V|}f_{\mathcal{G},r}(x) = \sum_{r=1}^{\infty} f_r^\circ(x).$$

We now combine this with (2.34), and use $\lim_{\mathcal{G}\to\mathbb{Z}^2}|E|/|V| = 2$ to obtain

$$-\beta f(\beta) = \lim_{\mathcal{G}\to\mathbb{Z}^2}\frac{1}{|V|}\ln Z^{\text{free}}_{\mathcal{G},\beta} = \ln(2\cosh^2\beta) + \sum_{r=1}^{\infty} f_r^\circ(x).$$

The low-temperature case can be treated in a similar manner, except that one must work on the dual graphs $\mathcal{G}^*$ with edge weights $x = \exp(-2\beta)$ on the dual edges, and use (2.31) instead of (2.33). □

Proof of Corollary 2.4. We will now show that Onsager's formula follows from the expressions for $f(\beta)$ derived above. First, we claim that with $x = \tanh\beta$ for $\beta \in (0,\beta_c)$ and $x = \exp(-2\beta)$ for $\beta \in (\beta_c,\infty)$, we have

$$-\beta f(\beta) = \ln\left[\frac{2\cosh 2\beta}{1+x^2}\right] + \sum_{r=1}^{\infty} f_r^\circ(x) \tag{2.35}$$

for all these β. This follows from (2.8) and the equality $\cosh^2\beta + \sinh^2\beta = \cosh^2\beta(1+x^2) = \cosh 2\beta$ for $\beta \in (0,\beta_c)$, and from (2.8) together with the logarithm of the equality $2\cosh 2\beta = e^{2\beta}(1+x^2)$ for $\beta \in (\beta_c,\infty)$.

In the proof of Theorem 2.3, we obtained $\sum_r f_r^\circ(x)$ as the limit of $|V|^{-1}\sum_r f_{\mathcal{G},r}(x)$. It is clear from the proof that here we may as well replace $f_{\mathcal{G},r}(x)$ by the corresponding sum of loop weights for the periodic graph $\mathcal{G}^{\circledcirc}$ from Section 2.3. In fact, the argument becomes even simpler on $\mathcal{G}^{\circledcirc}$, since we no longer have to treat vertices near the boundary separately. The transition matrix generating the loops in $\mathcal{G}^{\circledcirc}$ with the desired edge weights x is $x\Lambda^{\circledcirc}$. Hence, by Theorem 2.10 and the proof of Theorem 2.11 in Section 2.3, which gives $\|x\Lambda^{\circledcirc}\| \le x(\sqrt{2}+1)$, we see that for $x \in (0,\sqrt{2}-1)$,

$$\sum_{r=1}^{\infty} f_r^\circ(x) = \lim_{\mathcal{G}\to\mathbb{Z}^2}\frac{1}{|V|}\sum_{r=1}^{\infty} f_{\mathcal{G}^{\circledcirc},r}(x) = \lim_{\mathcal{G}\to\mathbb{Z}^2}\frac{1}{|V|}\frac{1}{2}\ln\det(\mathrm{I} - x\Lambda^{\circledcirc}). \tag{2.36}$$

We can compute $\det(\mathrm{I} - x\Lambda^{\circledcirc})$ by taking the Fourier transform of $\Lambda^{\circledcirc}$. This computation has appeared in the literature before, see for instance [27, 50, 55], but we also present it in brief form here for completeness.

Without loss of generality, we may assume that V is the set (2.29), in which case the Fourier transform of $\Lambda^{\circledcirc}$ is defined as

$$\widetilde{\Lambda}^{\circledcirc}_{(p,q)d,(p',q')d'} = \frac{1}{MN} \sum_{k,k'=0}^{M-1} \sum_{l,l'=0}^{N-1} e^{-\frac{2\pi i}{M}(pk-p'k') - \frac{2\pi i}{N}(ql-q'l')} \Lambda^{\circledcirc}_{(k,l)d,(k',l')d'},$$

where $d, d' \in \{\uparrow, \downarrow, \rightarrow, \leftarrow\}$. The calculation of this Fourier transform is made straightforward by the periodicity of $\Lambda^{\circledcirc}$, and reveals that the only entries surviving the summations are those for which $p' = p$ and $q' = q$. Hence, $\widetilde{\Lambda}^{\circledcirc}$ is a block-diagonal matrix of 4×4 blocks. To be precise, writing $\omega_p = 2\pi p/M$, $\omega_q = 2\pi q/N$, the 4×4 block for given p and q is

$$\widetilde{\Lambda}^{\circledcirc}_{(p,q)\cdot,(p,q)\cdot} = \begin{bmatrix} e^{i\omega_p} & e^{i\omega_p + i\pi/4} & 0 & e^{i\omega_p - i\pi/4} \\ e^{i\omega_q - i\pi/4} & e^{i\omega_q} & e^{i\omega_q + i\pi/4} & 0 \\ 0 & e^{-i\omega_p - i\pi/4} & e^{-i\omega_p} & e^{-i\omega_p + i\pi/4} \\ e^{-i\omega_q + i\pi/4} & 0 & e^{-i\omega_q - i\pi/4} & e^{-i\omega_q} \end{bmatrix}.$$

Since $\det(\mathrm{I} - x\Lambda^{\circledcirc}) = \det(\mathrm{I} - x\widetilde{\Lambda}^{\circledcirc})$, from this Fourier transform we obtain

$$\begin{aligned} \det(\mathrm{I} - x\Lambda^{\circledcirc}) &= \prod_{p=0}^{M-1} \prod_{q=0}^{N-1} \det\Big(\mathrm{I} - x\widetilde{\Lambda}^{\circledcirc}_{(p,q)\cdot,(p,q)\cdot}\Big) \\ &= \prod_{p=0}^{M-1} \prod_{q=0}^{N-1} \big[(1+x^2)^2 - 2x(1-x^2)(\cos\omega_p + \cos\omega_q)\big]. \end{aligned}$$

Using (2.36), we conclude that

$$\begin{aligned} \sum_{r=1}^{\infty} f_r^{\circ}(x) &= \lim_{M,N\to\infty} \frac{1}{2MN} \ln\det(\mathrm{I} - x\Lambda^{\circledcirc}) \\ &= \frac{1}{8\pi^2} \int_0^{2\pi}\!\!\int_0^{2\pi} \ln\big[(1+x^2)^2 - 2x(1-x^2)(\cos\omega_1 + \cos\omega_2)\big]\, d\omega_1\, d\omega_2. \end{aligned} \tag{2.37}$$

To finish the computation, note that by (2.35), we have

$$-\beta f(\beta) = \frac{1}{8\pi^2} \int_0^{2\pi}\!\!\int_0^{2\pi} \ln\left[\frac{4\cosh^2 2\beta}{(1+x^2)^2}\right] d\omega_1\, d\omega_2 + \sum_{r=1}^{\infty} f_r^{\circ}(x).$$

Combining this with (2.37), and then using the identity

$$\frac{2x(1-x^2)}{(1+x^2)^2} = \frac{\sinh 2\beta}{\cosh^2 2\beta},$$

which holds both for $x = \exp(-2\beta)$ and for $x = \tanh\beta$, we obtain

$$-\beta f(\beta) = \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{2\pi} \ln\bigl[4\cosh^2 2\beta - 4\sinh 2\beta(\cos\omega_1 + \cos\omega_2)\bigr]\, d\omega_1\, d\omega_2.$$

This is Onsager's formula for the isotropic Ising model on $\mathbb{Z}^2$. □

2.4.3 Low-temperature two-point functions

In this section, we discuss the Ising model with positive boundary conditions. We consider rectangles $\mathcal{G} = (V, E)$ in $\mathbb{Z}^2$ (which later tend to $\mathbb{Z}^2$) and denote by $\mathcal{G}^*$ the weak dual of $\mathcal{G}$. Recall that every spin configuration $\sigma \in \Omega_{\mathcal{G}}^+$ on $\mathcal{G}$ corresponds bijectively to an even subgraph of $\mathcal{G}^*$, that is, a graph in which all vertices in V^* have even degree. For given σ, we denote the corresponding even subset of E^* by $F(\sigma)$; for given even $F \subset E^*$, we denote the corresponding spin configuration by $\sigma(F)$.

Setting $x_e = e^{-2\beta}$ for every edge e in $\mathbb{Z}^{2*}$, by (2.30) and (2.31) we have

$$P_{\mathcal{G},\beta}^+(\sigma) = \frac{1}{Z_{\mathcal{G}^*}(x)} \prod_{e\in F(\sigma)} x_e, \qquad \sigma \in \Omega_{\mathcal{G}}^+, \tag{2.38}$$

where $Z_{\mathcal{G}^*}(x)$ is the generating function for $\mathcal{G}^*$ with edge weight vector $x = (x_e)_{e\in E^*}$. Note that here we implicitly restrict the edge weight vector x on $\mathbb{Z}^{2*}$ to the edges of the graph $\mathcal{G}^*$ we work on. Such implicit restrictions to the relevant edges will occur throughout this and the following sections.

Proof of Theorem 2.5. Fix $u, v \in \mathbb{Z}^2$, $u \neq v$, and let γ be a self-avoiding walk in $\mathbb{Z}^2$ from u to v. We may assume that $\mathcal{G}$ is large enough so that u, v and γ are all contained in the area spanned by $\mathcal{G}$, see Figure 2.1. We will express the two-point function $\langle\sigma_u\sigma_v\rangle_{\mathcal{G},\beta}^+$ as the quotient of two generating functions. To this end, we define new edge weights x'_e on the edges of $\mathbb{Z}^{2*}$ such that $x'_e = -x_e$ if e crosses γ, and $x'_e = x_e$ otherwise. The reason for defining the weights x'_e in this way is the crucial fact that for all $\sigma \in \Omega_{\mathcal{G}}^+$,

$$\sigma_u\sigma_v \prod_{e\in F(\sigma)} x_e = \prod_{e\in F(\sigma)} x'_e. \tag{2.39}$$

To see this, recall that the edges in $F(\sigma)$, by their very definition, cross edges $xy \in E$ for which $\sigma_x \neq \sigma_y$. If $\sigma_u = \sigma_v$, then following γ from u to v, we necessarily cross an even number of such edges. If $\sigma_u \neq \sigma_v$, then we cross an odd number. In either case, (2.39) holds.

With the help of (2.38), we can write

$$\langle\sigma_u\sigma_v\rangle_{\mathcal{G},\beta}^+ = \sum_{\sigma\in\Omega_{\mathcal{G}}^+} \sigma_u\sigma_v \mathbf{P}_{\mathcal{G},\beta}^+(\sigma) = \frac{1}{Z_{\mathcal{G}^*}(x)} \sum_{\sigma\in\Omega_{\mathcal{G}}^+} \sigma_u\sigma_v \prod_{e\in F(\sigma)} x_e,$$

and using (2.39) and the bijection between even F and $\Omega^+_{\mathcal{G}}$, we obtain

$$\langle \sigma_u \sigma_v \rangle^+_{\mathcal{G},\beta} = \frac{1}{Z_{\mathcal{G}^*}(x)} \sum_{\text{even } F \subset E^*} \prod_{e \in F} x'_e = \frac{Z_{\mathcal{G}^*}(x')}{Z_{\mathcal{G}^*}(x)}. \tag{2.40}$$

The idea that correlations in the Ising model can be studied by means of ratios of generating functions with changed edge weights (or equivalently, changed spin-spin interactions) has arisen before in the physics literature, see [37]. It now follows from Theorems 2.10 and 2.11 that for $\beta > \beta_c$, we have

$$\langle \sigma_u \sigma_v \rangle^+_{\mathcal{G},\beta} = \exp\biggl(\sum_{r=1}^{\infty} \sum_{\ell \in \mathcal{L}_r(\mathcal{G}^*)} \bigl[w(\ell; x') - w(\ell; x) \bigr] \biggr),$$

where $\mathcal{L}_r(\mathcal{G}^*)$ is the collection of loops of length r in the graph $\mathcal{G}^*$.

Recall that we call a loop in $\mathcal{G}^*$ *uv-odd* if it crosses γ an odd number of times. Observe that for uv-odd loops ℓ, $w(\ell; x') = -w(\ell; x)$, while for loops ℓ that are not uv-odd, $w(\ell; x') = w(\ell; x)$. It follows that

$$\langle \sigma_u \sigma_v \rangle^+_{\mathcal{G},\beta} = \exp\biggl(-2 \sum_{r=1}^{\infty} \sum_{\ell \in \mathcal{L}^{uv}_r(\mathcal{G}^*)} w(\ell; x) \biggr), \tag{2.41}$$

where $\mathcal{L}^{uv}_r(\mathcal{G}^*)$ is the collection of uv-odd loops of length r in the graph $\mathcal{G}^*$.

Note that a uv-odd loop of length r cannot travel far from u and v. To be precise, these loops must be contained in $B^u_r \cup B^v_r$, where B^u_r is a square in the plane of side length r centred at u, and B^v_r is defined similarly. To study the convergence of (2.41) as $\mathcal{G} \to \mathbb{Z}^2$, for arbitrary rectangles $\mathcal{G}$ in $\mathbb{Z}^2$ that can be finite or infinite, and even equal to $\mathbb{Z}^2$, we now define

$$a_r(\mathcal{G}^*; x) := \sum_{\ell \in \mathcal{L}^{uv}_r(\mathcal{G}^*)} w(\ell; x).$$

This definition makes sense both for finite and infinite $\mathcal{G}$, since the loops contributing to the sum must be contained in $B^u_r \cup B^v_r$.

Let B^{uv}_r denote the smallest rectangle in $\mathbb{R}^2$ containing both B^u_r and B^v_r, and write $\mathcal{G}^* \cap B^{uv}_r$ for the largest subgraph of $\mathcal{G}^*$ which is a rectangle in $\mathbb{Z}^{2*}$ entirely contained in B^{uv}_r. Then for all $\mathcal{G}$,

$$a_r(\mathcal{G}^*; x) = a_r(\mathcal{G}^* \cap B^{uv}_r; x) = \frac{1}{2} \sum_{\ell \in \mathcal{L}_r(\mathcal{G}^* \cap B^{uv}_r)} \bigl[w(\ell; x) - w(\ell; x') \bigr]. \tag{2.42}$$

Now, since the volume of B^{uv}_r is bounded from above by $(\|u-v\|+r)^2$, and $\exp(2\beta_c) = \sqrt{2}+1$ by (2.4), Theorem 2.11 yields the uniform bound

$$|a_r(\mathcal{G}^*; x)| \leq 2(\|u-v\|+r)^2 r^{-1} \exp\bigl(-2(\beta-\beta_c)r\bigr) \qquad \text{for all } \mathcal{G}. \tag{2.43}$$

We now return to (2.41). Since eventually, $\mathcal{G}^* \cap B^{uv}_r = \mathbb{Z}^{2*} \cap B^{uv}_r$ when $\mathcal{G} \to \mathbb{Z}^2$, from (2.42) we conclude that

$$a_r(\mathcal{G}^*; x) \to a_r\bigl(\mathbb{Z}^{2*}; x\bigr) \qquad \text{for all } r \geq 1.$$

Moreover, the $a_r(\mathcal{G}^*;x)$ are uniformly bounded in $\mathcal{G}$ by the right-hand side of (2.43), which is summable over r. Therefore, by dominated convergence,

$$\lim_{\mathcal{G}\to\mathbb{Z}^2}\sum_{r=1}^{\infty} a_r(\mathcal{G}^*;x) = \sum_{r=1}^{\infty} a_r\big(\mathbb{Z}^{2*};x\big),$$

where the series on the right is absolutely summable. Using (2.41), this proves the convergence of $\langle\sigma_u\sigma_v\rangle^+_{\mathcal{G},\beta}$ in Theorem 2.5.

Next, we consider $\langle\sigma_u\rangle^+_{\mathcal{G},\beta}$ for $u \in \mathcal{G}\setminus\partial\mathcal{G}$. We can treat this like $\langle\sigma_u\sigma_v\rangle^+_{\mathcal{G},\beta}$ by taking v on the boundary of $\mathcal{G}$, since then $\sigma_v = +1$. We now call a loop which crosses γ an odd number of times *u-odd*, since this depends only on u, not v. The box B_r^{uv} can be replaced by B_r^u in the argument, which replaces $(\|u-v\|+r)^2$ by r^2 in (2.43). This completes the proof of Theorem 2.5. □

Proof of Corollary 2.6. The limit $\langle\sigma_u\rangle^+_{\mathbb{Z}^2,\beta}$ in Theorem 2.5 is easily seen to be independent of the choice of u, and we take $u = o$ as the canonical choice. We now consider what happens to the two-point function when we take u and v further and further apart. When $r < \|u-v\|/2$, the boxes B_r^u and B_r^v in the proof of Theorem 2.5 above are disjoint. If this is the case, a uv-odd loop of length r in $\mathbb{Z}^{2*}$ must be contained in either B_r^u or B_r^v. Hence,

$$\begin{aligned}\langle\sigma_u\sigma_v\rangle^+_{\mathbb{Z}^2,\beta} = \exp&\bigg(-2\sum_{r\ge\|u-v\|/2} a_r\big(\mathbb{Z}^{2*};x\big)\bigg)\\ &\times\exp\bigg(-2\sum_{r<\|u-v\|/2}\Big[a_r(\mathbb{Z}^{2*}\cap B_r^u;x)+a_r\big(\mathbb{Z}^{2*}\cap B_r^v;x\big)\Big]\bigg).\end{aligned}$$

When $\|u-v\|\to\infty$, the first factor converges to 1 exponentially fast, since the uniform bound in (2.43) applies to $a_r\big(\mathbb{Z}^{2*};x\big)$. In the second factor, the first term in the sum is a sum over the u-odd loops of length r, and the second term is a sum over the v-odd loops. Hence the second factor factorizes and converges (exponentially fast) to $[\langle\sigma_o\rangle^+_{\mathbb{Z}^2,\beta}]^2$. □

2.4.4 High-temperature two-point functions

In this section, we discuss the Ising model on rectangles $\mathcal{G} = (V,E)$ in $\mathbb{Z}^2$ (which will again tend to $\mathbb{Z}^2$) with free boundary conditions. From the definitions (2.1), (2.2) and (2.3), we have

$$\langle\sigma_u\sigma_v\rangle^{\text{free}}_{\mathcal{G},\beta} = \sum_{\sigma\in\Omega^{\text{free}}_{\mathcal{G}}}\sigma_u\sigma_v\mathbf{P}^{\text{free}}_{\mathcal{G},\beta}(\sigma) = \frac{1}{\mathcal{Z}^{\text{free}}_{\mathcal{G},\beta}}\sum_{\sigma\in\Omega^{\text{free}}_{\mathcal{G}}}\sigma_u\sigma_v\prod_{wz\in E}e^{\beta\sigma_w\sigma_z}.$$

Performing the high-temperature expansion on the right-hand side of this expression, in the way explained in Section 2.4.1, leads to

$$\langle \sigma_u \sigma_v \rangle_{\mathcal{G},\beta}^{\text{free}} = \frac{2^{|V|}(\cosh\beta)^{|E|}}{\mathcal{Z}_{\mathcal{G},\beta}^{\text{free}}} \sum_{\substack{F\subset E\\ \delta F=\{u,v\}}} \prod_{e\in F} x_e,$$

where $x_e = \tanh\beta$ for every edge e in $\mathbb{Z}^2$, and δF denotes the set of all vertices that have odd degree in (V, F). Using (2.33), we conclude that

$$\langle \sigma_u \sigma_v \rangle_{\mathcal{G},\beta}^{\text{free}} = \frac{1}{Z_{\mathcal{G}}(x)} \sum_{\substack{F\subset E\\ \delta F=\{u,v\}}} \prod_{e\in F} x_e, \tag{2.44}$$

where $Z_{\mathcal{G}}(x)$ is the graph generating function for the graph $\mathcal{G}$ with edge weight vector $x = (x_e)_{e\in E}$.

Proof of Theorem 2.7. Fix $u, v \in \mathbb{Z}^2$, $u \neq v$, and recall the definitions of u^*, v^*, the walk γ and the additional edges E_γ and vertices V_γ from Section 2.1.2 (see Figure 2.1, right). For an arbitrary rectangle $\mathcal{G}$ in $\mathbb{Z}^2$ (either finite or infinite) containing u and v, we denote by $\mathcal{G}_\gamma$ the graph obtained from $\mathcal{G}$ by adding all vertices in V_γ to its vertex set, and all edges in E_γ to its edge set. In $\mathcal{G}_\gamma$, all edges added from the set E_γ are considered as additional, and the edges from $\mathcal{G}$ are considered as representative.

As in the low-temperature case, we now define weights x'_e on the edge set of $\mathbb{Z}^2$ such that $x'_e = -x_e$ if e crosses γ, and $x'_e = x_e$ otherwise. We also define edge weights $x'_\gamma(t)$ on the edge set of $\mathbb{Z}^2_\gamma$, as follows:

$$x'_\gamma(t)_e = \begin{cases} x'_e & \text{if } e \text{ is an edge of } \mathbb{Z}^2; \\ 1 & \text{if } e \in E_\gamma \setminus \{uu^*\}; \\ t & \text{if } e = uu^*. \end{cases} \tag{2.45}$$

To motivate this definition, consider a given rectangle $\mathcal{G} = (V, E)$ in $\mathbb{Z}^2$, large enough so that $u, v \in V$. We claim that

$$\sum_{\substack{F\subset E\\ \delta F=\{u,v\}}} \prod_{e\in F} x_e = \sum_{\substack{E_\gamma\subset F\subset E\cup E_\gamma\\ F \text{ even}}} (-1)^{C_F} \prod_{e\in F} x'_\gamma(1)_e. \tag{2.46}$$

To see this, note that we can bijectively map every F contributing to the first sum to a subgraph in the second sum, by taking the union $F \cup E_\gamma$. Doing this may introduce edge crossings, whence the factor $(-1)^{C_F}$, but these are compensated by switching from the edge weight vector x to $x'_\gamma(1)$.

The crucial step is now to recognize the last expression as the derivative of a graph generating function. Indeed, a simple consideration shows that

$$\sum_{\substack{E_\gamma\subset F\subset E\cup E_\gamma\\ F \text{ even}}} (-1)^{C(F)} \prod_{e\in F} x'_\gamma(1)_e = \frac{\partial}{\partial t}\Bigg(\sum_{\text{even } F\subset E\cup E_\gamma} (-1)^{C(F)} \prod_{e\in F} x'_\gamma(t)_e \Bigg),$$

evaluated at any t, since any even $F \subset E_\gamma$ contributes at most one factor t to the product of edge weights on the right. In particular, we are allowed to evaluate the derivative at $t = 0$. By (2.44) and (2.46), this establishes that

$$Z_{\mathcal{G}}(x) \cdot \langle \sigma_u \sigma_v \rangle^{\text{free}}_{\mathcal{G},\beta} = \frac{\partial}{\partial t} Z_{\mathcal{G}_\gamma}\bigl(x'_\gamma(t)\bigr)\Big|_{t=0}. \tag{2.47}$$

We now fix $\beta \in (0, \beta_c)$, so that by (2.4), $x_e \in (0, \sqrt{2} - 1)$ for every $e \in E$, and by Theorem 2.11, the spectral radius of $\Lambda(x')$ is strictly less than 1. We now need a similar bound on the spectral radius of the matrix $\Lambda_\gamma\bigl(x'_\gamma(t)\bigr)$, which is the transition matrix for the modified graph $\mathcal{G}_\gamma$ with edge weight vector $x'_\gamma(t)$. The difference between these two matrices is that $\Lambda_\gamma\bigl(x'_\gamma(t)\bigr)$ allows transitions between u and v along the chain of additional edges in E_γ. This means that the 32 matrix entries from du to vd' and from dv to ud', with $d, d' \in \{\uparrow, \downarrow, \rightarrow, \leftarrow\}$, are nonzero in $\Lambda_\gamma\bigl(x'_\gamma(t)\bigr)$ for $t \neq 0$, while they are 0 in $\Lambda(x')$; all other entries of the two matrices are the same.

The 32 deviating matrix entries are all of the form $te^{i\phi/2}$, where ϕ is a sum of turning angles. Here, t will be treated as a complex variable. For $t = 0$, $\Lambda_\gamma\bigl(x'_\gamma(t)\bigr) = \Lambda(x')$. Since the eigenvalues vary continuously with t, we conclude that there exists $\varepsilon > 0$ such that for all t satisfying $|t| < \varepsilon$, the spectral radius of $\Lambda_\gamma\bigl(x'_\gamma(t)\bigr)$ is bounded from above by some $\alpha \in (0, 1)$.

Hence, if $|t| < \varepsilon$, Theorem 2.10 applies, and we obtain

$$Z_{\mathcal{G}_\gamma}\bigl(x'_\gamma(t)\bigr) = \exp\biggl(\sum_{r=1}^{\infty} f_{\gamma r}(t)\biggr),$$

where

$$f_{\gamma r}(t) = \sum_{\ell \in \mathcal{L}_r(\mathcal{G}_\gamma)} w\bigl(\ell; x'_\gamma(t)\bigr).$$

Note that, this last sum being finite, the $f_{\gamma r}(t)$ are polynomials in t. Also, from (2.21) it follows that $|f_{\gamma r}(t)| \leq 2|V|\alpha^r$. Therefore, the partial sums of the series $\sum_r f_{\gamma r}(t)$ are uniformly convergent for $|t| < \varepsilon$, and the sum of the series is an analytic function of t. Moreover, the derivatives of the partial sums also converge uniformly to the derivative of the sum of the series.

From all this, it follows that the right-hand side of (2.47) is equal to

$$\biggl(\sum_{r=1}^{\infty} \frac{\partial}{\partial t} \sum_{\ell \in \mathcal{L}_r(\mathcal{G}_\gamma)} w\bigl(\ell; x'_\gamma(t)\bigr)\Big|_{t=0}\biggr) \exp\biggl(\sum_{r=1}^{\infty} \sum_{\ell \in \mathcal{L}_r(\mathcal{G}_\gamma)} w\bigl(\ell; x'_\gamma(0)\bigr)\biggr).$$

In the first factor, the only loops that survive the differentiation are those that visit the edge uu^*, since only they contribute a factor t to the weight. Taking the derivative at $t = 0$, we are only left with those loops that visit the edge uu^* exactly once. In the second factor, because we set $t = 0$, the only loops that contribute are those that do not visit uu^*. This leaves precisely all loops in the graph $\mathcal{G}$. The right-hand side of (2.47) therefore becomes

$$\left(\sum_{r=1}^{\infty}\sum_{\ell\in\mathcal{L}_r^{uu^*}(\mathcal{G}_\gamma)} w\big(\ell;x'_\gamma(1)\big)\right)\exp\left(\sum_{r=1}^{\infty}\sum_{\ell\in\mathcal{L}_r(\mathcal{G})} w(\ell;x')\right), \tag{2.48}$$

where $\mathcal{L}_r^{uu^*}(\mathcal{G}_\gamma)$ is the set of loops of length r in $\mathcal{G}_\gamma$ that visit uu^* once. From this, applying Theorem 2.10 again to the second factor, we find that

$$\langle\sigma_u\sigma_v\rangle_{\mathcal{G},\beta}^{\text{free}} = \left(\sum_{r=1}^{\infty}\sum_{\ell\in\mathcal{L}_r^{uu^*}(\mathcal{G}_\gamma)} w\big(\ell;x'_\gamma(1)\big)\right)\frac{Z_{\mathcal{G}}(x')}{Z_{\mathcal{G}}(x)}. \tag{2.49}$$

Note the ratio of graph generating functions in (2.49). Recall that we have seen such a ratio of graph generating functions before in the low-temperature case, namely in (2.40). Thus, this ratio can be interpreted as a two-point function between the spins at u^* and v^* in a dual Ising model with positive boundary conditions at the dual low temperature β^*, given by $\exp(-2\beta^*) = \tanh\beta$. Using (2.41), we can express this ratio in terms of a sum over all u^*v^*-odd loops in the graph $\mathcal{G}$, if we like.

Next, we want to consider the limit as $\mathcal{G}\to\mathbb{Z}^2$. By the argument given in Section 2.4.3, we already know that the ratio of graph generating functions in (2.49) converges to $\langle\sigma_{u^*}\sigma_{v^*}\rangle^{+}_{\mathbb{Z}^{2*},\beta^*}$. It remains to consider what happens to the sum over the loops that visit uu^* once. To this end, for a general finite or infinite rectangle $\mathcal{G}$ in $\mathbb{Z}^2$ containing u and v, we define

$$a_r(\mathcal{G}_\gamma;x'_\gamma) := \sum_{\ell\in\mathcal{L}_r^{uu^*}(\mathcal{G}_\gamma)} w(\ell;x'_\gamma), \tag{2.50}$$

where we have simplified the notation by letting $x'_\gamma \equiv x'_\gamma(1)$.

As in the low-temperature case, the loops that contribute to $a_r(\mathcal{G}_\gamma;x'_\gamma)$ must be confined to the box B_r^{uv}, defined in the same way as before, except possibly for the steps taken along the additional edges in E_γ, which are not counted in the length of the loop, and are allowed to go outside B_r^{uv}. Hence,

$$a_r(\mathcal{G}_\gamma;x'_\gamma) = a_r\big((\mathcal{G}\cap B_r^{uv})_\gamma;x'_\gamma\big), \tag{2.51}$$

where $\mathcal{G}\cap B_r^{uv}$ is the largest subgraph of $\mathcal{G}$ contained in B_r^{uv}, as before. It follows that $a_r(\mathcal{G}_\gamma;x'_\gamma)\to a_r\big(\mathbb{Z}^2_\gamma;x'_\gamma\big)$ for all $r\geq 1$. As before, we now want to use dominated convergence to prove that

$$\lim_{\mathcal{G}\to\mathbb{Z}^2}\sum_{r=1}^{\infty} a_r(\mathcal{G}_\gamma;x'_\gamma) = \sum_{r=1}^{\infty} a_r\big(\mathbb{Z}^2_\gamma;x'_\gamma\big), \tag{2.52}$$

and that the right-hand side is absolutely summable. This requires an appropriate uniform bound (in $\mathcal{G}$) on the right-hand side of (2.50).

To obtain this bound, by (2.51) it is sufficient to consider an arbitrary finite rectangle $\mathcal{G}$ in $\mathbb{Z}^2$ containing u and v, and such that $\mathcal{G}$ is contained in B_r^{uv}. Note that every loop of length r in $\mathcal{G}_\gamma$ which visits the edge uu^* once, has a representation of the form $\omega \oplus \gamma$, where ω is a walk of length r in $\mathcal{G}$ from v to u. Here, we use the facts that the part ω of the loop never visits uu^*, and that the steps taken along γ from u to v do not contribute towards the length of the loop, since they are along additional edges.

Let $\Lambda_R(x')$ be the transition matrix for the graph $\mathcal{G}$ with edge weights x'_e. Note that the sum of the weights of all walks ω of length r from $v{\rightarrow}$ to $\uparrow u$, for instance, is given by the entry of the matrix $\Lambda_R^r(x')$ in row $v{\rightarrow}$ and column $\uparrow u$. To compute the sum of the weights of the corresponding loops $\omega \oplus \gamma$, we only need to multiply this entry by the factor $e^{i\phi/2}$, where ϕ is the sum of the turning angles encountered in the walk from $\uparrow u$ to $v{\rightarrow}$ along the edges in E_γ. From these observations, we can conclude that

$$a_r(\mathcal{G}_\gamma; x'_\gamma) \leq 16\, \|\Lambda_R^r(x')\|_{\max},$$

where $\|\cdot\|_{\max}$ denotes the maximum-entry norm, and the factor 16 comes from the fact that there are 4 directed (representative) edges pointing out from v, and 4 pointing to u. By Theorem 2.11 and the fact that the maximum-entry norm of a matrix is bounded by the operator norm, and using that $(\tanh\beta_c)^{-1} = \sqrt{2}+1$ by (2.4), we obtain

$$a_r\big(\mathcal{G}_\gamma; x'_\gamma\big) \leq 16\, \|\Lambda_R^r(x')\| \leq 16\, \|\Lambda_R(x')\|^r \leq 16 \left(\frac{\tanh\beta}{\tanh\beta_c}\right)^r. \tag{2.53}$$

We emphasize that this bound holds uniformly for all finite and infinite rectangles $\mathcal{G}$ containing u and v. Hence, (2.52) holds by dominated convergence, and this completes the proof of Theorem 2.7. □

Proof of Corollary 2.8. We now consider what happens to the two-point function studied above when we let $\|u-v\|$ tend to infinity. Since the loops that visit uu^* necessarily have length at least $\|u-v\|$, we can write

$$\langle\sigma_u\sigma_v\rangle^{\text{free}}_{\mathbb{Z}^2,\beta} = \left(\sum_{r\geq\|u-v\|} a_r\big(\mathbb{Z}^2_\gamma; x'_\gamma\big)\right)\langle\sigma_{u^*}\sigma_{v^*}\rangle^{+}_{\mathbb{Z}^{2*},\beta^*}.$$

By Theorem 2.5, the two-point function on the right is bounded between 0 and 1. Alternatively, at this stage we could also observe that the ratio of graph generating functions in (2.49) is always between -1 and $+1$, since the same even subgraphs contribute to both generating functions, but only in the numerator, some of them come with a negative sign. Furthermore, the bound in (2.53) holds for $a_r\big(\mathbb{Z}^2_\gamma; x'_\gamma\big)$. This gives the desired upper bound. That the two-point function is nonnegative follows directly from (2.44).□