Heisenberg Model, Bethe *Ansatz*, and Random Walks

Lenhard L. Ng
Senior Honors Thesis
Harvard University
1 April 1996

Prof. Persi Diaconis, advisor

in partial fulfillment of the honors requirements
for the degree of
Bachelor of Arts
in Mathematics and Physics

261 Quincy Mail Center
Cambridge, MA 02138
(617) 493-7613

Contents

1 Introduction

Consider the following model for card shuffling: a group of N cards, each colored red or black, is arranged in a ring, and the shuffler repeatedly chooses two adjacent cards and switches them. This simple model, which we call the interchange model, is an example of a Markov chain; an in-depth analysis of this chain has never been done, and one of our goals in this paper is to accomplish this analysis. The interchange model is in fact essentially equivalent to a special case of another Markov chain known as the exclusion process, which has wide applications to mathematical models in physics and biology, and so results about the interchange chain produce new results about the exclusion process as well.

One of the most remarkable and serendipitous properties of the interchange model is that the transition matrix describing it is precisely the matrix corresponding to a Hamiltonian operator in physics, the so-called Heisenberg XXX Hamiltonian, describing a quantum mechanical system of spin-1/2 particles. In this system, there are N particles arranged in a ring, and each particle possesses a spin of either up or down; neighboring particles interact via so-called spin-spin interactions, through which the energy associated to them depends on whether their spins point in the same direction or in opposite directions. A generalized, three-dimensional version of this model has been experimentally shown to be an accurate model for the phenomena of ferromagnetism and antiferromagnetism, and some materials (e.g., $YFeO_3$) have even been shown to behave according to the somewhat simplistic one-dimensional model described here. Following its discovery in 1926, the Heisenberg model attracted much attention in the physics community, and, in 1931, Hans Bethe formulated the famous Bethe *ansatz* to approach the one-dimensional Heisenberg model. Physicists have since been able to use the Bethe *ansatz* to find all of the energy levels of the one-dimensional Heisenberg system.

The correspondence between the interchange process and the Heisenberg model is simply a correspondence between cards and spins; there are two types of both cards and spins, and, in each case, the interaction is only between nearest neighbors in a ring. What is astonishing, and not at all obvious, is that the spin-spin interaction from the Heisenberg XXX model is mathematically equivalent to the card-shuffling "interaction" consisting of switching nearest neighbors. Through this equivalence, the methods used by physicists to find the energy levels of the Heisenberg model, and in particular the Bethe *ansatz*, can also be used by mathematicians to find the eigenvalues of the transition matrix associated to the interchange model; these eigenvalues, in turn, give information about how quickly the interchange shuffle becomes random. The Bethe *ansatz* approach from physics can thus be applied to mathematical random walk theory to yield results, difficult to verify by other means, about the interchange walk and other Markov chains as well.

We now outline the contents of this paper. In Section 2, we give the necessary background from physics, by describing, deriving, and analyzing the Heisenberg model, paying special attention to the one-dimensional XXX case. Section 3 is a diagonalization of the interchange transition matrix, using the Bethe *ansatz*, in the cases in which there are either one or two red cards. In Section 4, we examine consequences of this diagonalization, culminating in the calculation in Section 5 of the rate at which the interchange process becomes random.

Section 6 is devoted to other, related random walks, including the exclusion process and the general interchange process; it concludes with an observation through which estimates from random walk theory can be used in thermodynamics.

We will assume familiarity with some amount of analysis and random walk theory, although we review elementary results about random walks in Section 5.1; in Section 6.2, we also use a bit of representation theory. We do not assume familiarity with quantum mechanics, although our review of quantum mechanical theory in Section 2 is designed to be unobtrusive for those who do have some background in quantum mechanics.

Much of this paper is, we believe, original, since the approach of applying physics to the interchange model, although hinted at by several sources, has apparently not yet been carried out in a way that uses results from random walk theory. Section 2 is mainly expository; Section 3 reformulates results already known to Bethe in 1931, but is original in its use of a mathematical level of rigor rather than a physical trust of intuition. Sections 4, 5, and 6 are more or less entirely original, except where noted otherwise.

2 Physical background

This section is almost purely physics, as we give the background necessary to understand the remainder of the paper. We begin with an explanation in Section 2.1 of the importance of the Heisenberg model to physics, and then proceed to define the Heisenberg model in mathematical terms in Section 2.2. We then give a physical derivation in Section 2.3 of the Heisenberg model, and discuss physical approaches to solving the Heisenberg model in Section 2.4.

2.1 The Heisenberg model of magnetism

The Heisenberg model occupies a somewhat uncomfortable position in solid state and mathematical physics. Easy to state and widely applicable, it is also extremely difficult to analyze. The interaction model which became known as the Heisenberg model was discovered by Werner Heisenberg and, nearly simultaneously, by P. A. M. Dirac, in 1926. Further work over the following decade established the Heisenberg model, due to its versatility, to be "the fundamental object of study of the theory of magnetism" [13, p. 40]. Since then, much work in theoretical statistical mechanics, solid state physics, and mathematical physics alike has been concentrated on understanding this model. Despite this effort, the Heisenberg model remains largely intractable, its predicted energy levels and eigenfunctions poorly understood, especially in a three-dimensional setting.

The first breakthrough towards solving the Heisenberg model was the diagonalization by Hans Bethe [2] in 1931 of the one-dimensional Heisenberg XXX model; the *ansatz*, or hypothesis, formulated in the process was later applied to many other physical models. Using an extension of the Bethe *ansatz*, Rodney Baxter solved the more general one-dimensional XYZ model in 1971. This solution, however, as well as approaches to the two-dimensional and three-dimensional Heisenberg models, is exceedingly complicated. The one-dimensional XXX model, and its solution, are still of interest, perhaps because they are somewhat well understood; fortunately, we will only require analysis of this model.

The Heisenberg model can be used to describe magnetically ordered solids, in which internal magnetic interactions cause individual magnetic ions to possess nonzero magnetic moments below some critical temperature. Magnetically ordered solids separate broadly into three groups: ferromagnets, antiferromagnets, and ferrimagnets. In ferromagnets, on which we will concentrate, the local moments are energetically preferred to be all aligned in a particular direction, the direction of spontaneous magnetization, so that the solid as a whole has a nonzero spontaneous magnetic moment. In antiferromagnets, adjacent local moments prefer to be antialigned; ferrimagnets have a nonzero spontaneous magnetization, but not necessarily aligned local moments.

For those who are unfamiliar with quantum mechanics, we conclude this section by introducing some quantum mechanical notation which will be used in subsequent sections. Quantum mechanics postulates that all information about the state of a system is given by that state's wave function, which is an element ψ, usually of norm 1, in a Hilbert space $\mathfrak{H}$ over $\mathbb{C}$. Typically, $\mathfrak{H} = L^2(K)$ for some $K \subset \mathbb{R}^n$, with the usual complex inner product; in that case, $|\psi(\mathbf{r})|^2$ gives the probability density of finding the system at point $\mathbf{r}$. The behavior

of the system is controlled by the Hamiltonian $\mathcal{H}$, a Hermitian operator on the Hilbert space in question, via another postulate of quantum mechanics, the time-independent Schrödinger equation $\mathcal{H}\psi = E\psi$. (For a one-particle system, $\mathcal{H}(x)$ is typically of the form $-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})$, where m is the mass of the particle, ∇^2 is the Laplacian operator, and $V(\mathbf{r})$ is the potential experienced by the particle.) The wave function ψ is then an eigenvector of $\mathcal{H}$, and the corresponding eigenvalue E is the energy of the system. By the spectral theorem, any hermitian operator is diagonalizable, with all real eigenvalues; in particular, any energy eigenvalue is real. In the following sections, we will use several times the fact that any two commuting hermitian operators can be diagonalized by a basis of simultaneous eigenvectors. In accordance with most physics textbooks, we will use Dirac notation when describing spin states; thus $|v\rangle$ represents the vector described by v.

2.2 Mathematical formulation of the one-dimensional Heisenberg model $\mathcal{H}$

In this section, we formulate the Heisenberg Hamiltonian $\mathcal{H}$ in mathematical language, and make some preliminary remarks. The system under study is a one-dimensional system of N interacting spin-1/2 fermions, with interactions described by the Heisenberg Hamiltonian; we will follow the notation of Takhtadzhan and Faddeev [16]. In mathematical terms, $\mathcal{H}$ acts on the 2^N–dimensional tensor product of N spin spaces $\mathfrak{H} = \bigotimes_{j=1}^N \mathfrak{h}_j$. Here $\mathfrak{h}_j$ is a two-dimensional vector space over $\mathbb{C}$; the standard basis $e_j^+ = \binom{1}{0}$, $e_j^- = \binom{0}{1}$ of $\mathfrak{h}_j$ corresponds in physics to the two possible spin states $|\uparrow\rangle$, $|\downarrow\rangle$ of a spin-1/2 fermion. In defining operators on $\mathfrak{H}$ in terms of operators on the $\mathfrak{h}_j$, we will use the following convenient notation: if ρ is a linear operator on $\mathfrak{h}_j$, then denote by ρ_j the operator on $\mathfrak{H}$ defined by

$$\rho_j = \underbrace{1 \otimes \cdots \otimes 1}_{j-1} \otimes \rho \otimes 1 \otimes \cdots \otimes 1.$$

Intuitively, ρ_j acts precisely on the j-th component of the tensor product.

We define operators $\sigma^x, \sigma^y, \sigma^z$ on any of the spaces $\mathfrak{h}_j$ by the usual Pauli spin matrices

$$\sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

where the matrices are with respect to the standard basis e_j^+, e_j^-. These spin matrices give operators σ_j^α on $\mathfrak{H}$. In terms of the Pauli spin operators, the Heisenberg XYZ Hamiltonian is given by

$$\mathcal{H} = -\frac{1}{2}\sum_{j=1}^{N}(J_x\sigma_j^x\sigma_{j+1}^x + J_y\sigma_j^y\sigma_{j+1}^y + J_z\sigma_j^z\sigma_{j+1}^z). \tag{2.1}$$

Here σ_{N+1}^α is understood to mean σ_1^α, and J_x, J_y, J_z are real constants. Note that $\mathcal{H}$ is real symmetric and therefore hermitian.

P. A. M. Dirac [6] observed that, when $J_x = J_y = J_z = J$ (the so-called XXX model), the above Hamiltonian operator has an alternative interpretation involving "switching spins."

In accordance with the notation in [15], let the operator $P_{j,k}$ act on $\mathfrak{H}$ by permuting $\mathfrak{h}_j$ and $\mathfrak{h}_k$; for instance, $P_{2,3}$ maps $e_1^- \otimes e_2^+ \otimes e_3^-$ to $e_1^- \otimes e_2^- \otimes e_3^+$, and $e_1^- \otimes e_2^+ \otimes e_3^+$ to itself. We next introduce the usual spin raising and lowering operators $\sigma^\pm = (\sigma^x \pm i\sigma^y)/2$ on $\mathfrak{h}_j$, which act on the generators of $\mathfrak{h}_j$ by $\sigma^+(e^+) = 0$, $\sigma^+(e^-) = e^+$, $\sigma^-(e^+) = e^-$, $\sigma^-(e^-) = 0$. Then it is easy to see that

$$P_{j,j+1} = \frac{\sigma_j^z \sigma_{j+1}^z + 1}{2} + \sigma_j^+ \sigma_{j+1}^- + \sigma_j^- \sigma_{j+1}^+ = \frac{1}{2}(1 + \sigma_j^x \sigma_{j+1}^x + \sigma_j^y \sigma_{j+1}^y + \sigma_j^z \sigma_{j+1}^z),$$

so that the XXX Hamiltonian defined by (2.1) can be written as

$$\mathcal{H} = \frac{N}{2} - J \sum_{j=1}^{N} P_{j,j+1}. \tag{2.2}$$

This equation is the crucial step in linking the Heisenberg Hamiltonian to a Markov chain involving interchanging neighboring cards in a circle.

Note that (2.2) implies that the XXX Hamiltonian conserves the numbers of e^+ and e^- in any basis element of $\mathfrak{H}$, since it merely permutes them. This conservation holds more generally if we merely assume that $J_x = J_y$, the so-called XXZ model. Indeed, when $J_x = J_y$, we may rewrite $\mathcal{H}$ as

$$\mathcal{H} = -\frac{1}{2} \sum_{j=1}^{N} (2J_x(\sigma_j^+ \sigma_{j+1}^- + \sigma_j^- \sigma_{j+1}^+) + J_z \sigma_j^z \sigma_{j+1}^z). \tag{2.3}$$

It is then apparent that $\mathcal{H}$ conserves the number of e^+, and the number of e^-, appearing in a basis element of $\mathfrak{H}$. More precisely, introduce the operator $\sigma_{\text{tot}}^z = \sum_{j=1}^N \sigma_j^z$, the sum total z spin of the system. (Any member of the standard basis of $\mathfrak{H}$ is an eigenvector of σ_{tot}^z, with eigenvalue equal to the difference between the number of e^+ and e^- in its representation. For instance, if $N = 3$ and $\psi = |\downarrow\uparrow\downarrow\rangle = e_1^- \otimes e_2^+ \otimes e_3^-$, then $\sigma_j^z(\psi) = (1-2)\psi = -\psi$, since there are one e^+ and two e^- in ψ.)

Each of $\sigma_j^+ \sigma_{j+1}^-$, $\sigma_j^- \sigma_{j+1}^+$, and $\sigma_j^z \sigma_{j+1}^z$ conserves total z spin (to reiterate, this is intuitively $\#(e^+) - \#(e^-)$) for the standard basis of $\mathfrak{H}$, and so $\mathcal{H}$ commutes with σ_j^z on $\mathfrak{H}$. Thus the matrix for $\mathcal{H}$ splits into sections corresponding to the various possible values of σ_{tot}^z. (In physical language, σ_{tot}^z is a "good quantum number," so that we can find simultaneous eigenvectors of $\mathcal{H}$ and σ_{tot}^z.) If we denote by $\mathfrak{H}_r$ the $\binom{N}{r}$-dimensional subspace of $\mathfrak{H}$ consisting of eigenvectors of σ_{tot}^z with eigenvalue $N - 2r$ (that is, linear combinations of the $\binom{N}{r}$ configurations of r down spins and $N - r$ up spins), then $\mathfrak{H} = \mathfrak{H}_0 \oplus \mathfrak{H}_1 \oplus \mathfrak{H}_2 \oplus \cdots \oplus \mathfrak{H}_N$, and $\mathcal{H}$ acts on each of the $\mathfrak{H}_r$ separately; denote the restriction of $\mathcal{H}$ to $\mathfrak{H}_r$ by $\mathcal{H}_r$, and observe that the matrix representation of $\mathcal{H}_r$ is of size $\binom{N}{r} \times \binom{N}{r}$.

In the language of Markov chains, $\mathcal{H}_r$ roughly corresponds, in the XXX case, to the transition matrix of the interchange chain on r red and $N - r$ black cards; see Section 4.1 for further details. Since the steps of the Markov chain preserve the number of red cards (or, equivalently, the number of down spins), examining the action of $\mathcal{H}_r$ on $\mathfrak{H}_r$ corresponds to looking at the Markov chain on all configurations of r red and $N - r$ black cards, or on r-subsets of an N-set.

For future reference, we will use the following notation. The standard basis of $\mathfrak{H}_r$ is the collection of $\binom{N}{r}$ vectors

$$\{|x_1 x_2 \cdots x_r\rangle : 1 \leq x_1 < x_2 < \cdots < x_r \leq N\},$$

where $x_1, x_2, \ldots, x_r$ are integers, and $|x_1 x_2 \cdots x_r\rangle$ is the tensor product of the vectors e^- in positions $x_1, x_2, \ldots, x_r$ and e^+ everywhere else; for instance, when $N = 5$ and $m = 2$, then $|3, 5\rangle = e^+ \otimes e^+ \otimes e^- \otimes e^+ \otimes e^-$.

In Section 3, we will diagonalize $\mathcal{H}_r$ when $r = 1$ and $r = 2$; here the Bethe *ansatz* will prove to be crucial. First, however, we backtrack a bit and derive the Heisenberg model from a physical standpoint.

2.3 Physical derivation of the general Heisenberg model

In this section, we formulate, and provide the motivation behind, the general Heisenberg model of magnetism, in the language of physics. As we go along, we will give some amount of background in quantum mechanics for those unfamiliar with the basic theory. Please note that this section exists primarily to motivate the study of the Heisenberg model, and, as such, is mostly unrelated to the remainder of the paper; it may be skipped without greatly affecting understanding of the rest of the paper. The physics in this section is very loosely based on Keffer [11].

We derive the Heisenberg model beginning with a two-electron system under an electrostatic Coulomb interaction potential. Electron j has a position in $\mathbb{R}^3$ as well as a spin (up or down, or some combination), so that its wave function is an element of $\mathbf{H}_j \cong L^2(\mathbb{R}^3) \otimes \mathfrak{h}_j$, the tensor product of spatial and spin wave function spaces. (Here $\mathfrak{h}_j \cong \mathbb{C}^2$ is a two-dimensional vector space generated by $e_j^+ = |\uparrow\rangle$ and $e_j^- = |\downarrow\rangle$, as in Section 2.2.) The total wave function is then an element of $\mathbf{H}_1 \otimes \mathbf{H}_2 = \mathfrak{H}_{\text{spatial}} \otimes \mathfrak{H}_{\text{spin}}$, where $\mathfrak{H}_{\text{spatial}} \cong L^2(\mathbb{R}^3) \otimes L^2(\mathbb{R}^3)$ is the spatial component of $\mathbf{H}_1 \otimes \mathbf{H}_2$, and $\mathfrak{H}_{\text{spin}} \cong \mathfrak{h}_1 \otimes \mathfrak{h}_2$ is the spin component of $\mathbf{H}_1 \otimes \mathbf{H}_2$. The key observation here is that the Hamiltonian describing the Coulombic interaction, which acts only on $\mathfrak{H}_{\text{spatial}}$, can be formally replaced by a "spin Hamiltonian" acting only on $\mathfrak{H}_{\text{spin}}$. We will now briefly describe the reasoning behind this replacement.

The Coulomb interaction Hamiltonian is given by

$$\mathcal{H}_{\text{Coul}} = -\frac{\hbar^2}{2m}(\nabla_1^2 + \nabla_2^2) + \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|},$$

where $\mathcal{H}_{\text{Coul}}$ acts on $\mathbf{H}_1 \otimes \mathbf{H}_2$, ∇^2 is the usual Laplacian operator on $L^2(\mathbb{R}^3)$ (or, more precisely, on $L^2(\mathbb{R}^3) \cap C^2(\mathbb{R}^3)$, but we will assume that all wave functions are twice differentiable), and $\frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|}$ represents the electrostatic Coulomb repulsion potential between the two electrons. (The exact form of this Hamiltonian is relatively unimportant, and has been included only for completeness.) Note that $\mathcal{H}$ depends only on the spatial coordinates $\mathbf{r}_1, \mathbf{r}_2$ of the electrons, and not on their spins, and therefore it essentially acts only on the spatial component $\mathfrak{H}_{\text{spatial}}$ of $\mathbf{H}_1 \otimes \mathbf{H}_2$.

We now invoke the Pauli exclusion principle, which postulates that the total wave function describing the two electrons be antisymmetric, i.e., that it is replaced by its negative

when $\mathbf{H}_1$ and $\mathbf{H}_2$ are formally interchanged. Since the total wave function must be an eigenvector of $\mathcal{H}$, it is easy to see that it is expressible as the tensor product $\psi_{\text{spatial}} \otimes \psi_{\text{spin}}$ of an eigenvector of $\mathcal{H}$ in $\mathfrak{H}_{\text{spatial}}$ and a vector in $\mathfrak{H}_{\text{spin}}$. Then one of ψ_{spatial} and ψ_{spin} is symmetric, and the other is antisymmetric. Symmetric and antisymmetric spatial wave functions are given by

$$\psi^{\text{sym}}_{\text{spatial}} = \psi_1(\mathbf{r}_1) \otimes \psi_2(\mathbf{r}_2) + \psi_2(\mathbf{r}_1) \otimes \psi_1(\mathbf{r}_2) \quad \text{and} \quad \psi^{\text{anti}}_{\text{spatial}} = \psi_1(\mathbf{r}_1) \otimes \psi_2(\mathbf{r}_2) - \psi_2(\mathbf{r}_1) \otimes \psi_1(\mathbf{r}_2),$$

where ψ_1 and ψ_2 are wave functions in $L^2(\mathbb{R}^3)$ determined by $\mathcal{H}_{\text{Coul}}$ and the time-independent Schrödinger equation.

Symmetric and antisymmetric spin wave functions, on the other hand, are given by

$$\psi^{\text{sym}}_{\text{spin}} = |\uparrow\uparrow\rangle,\ |\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle,\ \text{or}\ |\downarrow\downarrow\rangle, \quad \text{and} \quad \psi^{\text{anti}}_{\text{spin}} = |\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle.$$

(Here, e.g., $|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle$ represents $e_1^+ \otimes e_2^- - e_1^- \otimes e_2^+$; this is standard physical notation.) The symmetric and antisymmetric spin vectors may be distinguished by their eigenvalues with respect to the spin operator $\sigma_1 \cdot \sigma_2 = \sigma_1^x\sigma_2^x + \sigma_1^y\sigma_2^y + \sigma_1^z\sigma_2^z$, where σ_j^α is the operator on $\mathfrak{h}_j$ defined as in Section 2.2; it is straightforward to check that the symmetric spin vectors are eigenvectors of $\sigma_1 \cdot \sigma_2$ with eigenvalue 1, while the antisymmetric spin vector is an eigenvector of $\sigma_1 \cdot \sigma_2$ with eigenvalue -3.[1]

Under the Coulomb Hamiltonian $\mathcal{H}_{\text{Coul}}$, the two spatial wave functions $\psi^{\text{sym}}_{\text{spatial}}$ and $\psi^{\text{anti}}_{\text{spatial}}$ have different energies E^{sym} and E^{anti}, which may be calculated from the Hamiltonian. Although the origin of this energy difference is the spatial Coulombic interaction, we may recast it formally as a spin-spin interaction through our calculation of the eigenvalues of the ψ_{spin} vectors with respect to $\sigma_1 \cdot \sigma_2$. Indeed, the "spin Hamiltonian"

$$\mathcal{H}_{\text{spin}} = \frac{3E^{\text{sym}} + E^{\text{anti}}}{4} + \frac{E^{\text{sym}} - E^{\text{anti}}}{4}\sigma_1 \cdot \sigma_2$$

has eigenvalue E^{anti} for the triplet symmetric spin states (which correspond to spatial wave function $\psi^{\text{anti}}_{\text{spatial}}$) and E^{sym} for the singlet antisymmetric spin state (which corresponds to spatial wave function $\psi^{\text{sym}}_{\text{spatial}}$), and is thus formally equivalent to $\mathcal{H}_{\text{Coul}}$. Since energies are relative, we may ignore the constant term in the above expression, obtaining $\mathcal{H}_{\text{spin}} = -\frac{J}{2}\sigma_1 \cdot \sigma_2$, where J, the so-called exchange integral, is given by $J = (E^{\text{sym}} - E^{\text{anti}})/2$. The sign of J indicates whether aligned or antialigned spins are energetically preferred, so that $J > 0$ for a ferromagnet, while $J < 0$ for an antiferromagnet. (This does not affect the eigenvectors, but does affect which eigenvectors correspond to the "ground state" or "low-lying excitations"; for instance, the ground state of the antiferromagnet will be the highest

[1]For those familiar with quantum mechanics, there is a simple explanation for the eigenvalues of the triplet (symmetric) and singlet (antisymmetric) spin vectors with respect to $\sigma_1 \cdot \sigma_2$. If the spin angular momentum vectors of the two electrons are $\mathbf{S}_1 = (S_1^x, S_1^y, S_1^z)$ and $\mathbf{S}_2 = (S_2^x, S_2^y, S_2^z)$ (in units of $\hbar$), then $S_j^\alpha = \sigma_j^\alpha/2$. Now the triplet (total spin quantum number $S = 1$) and singlet ($S = 0$) states are eigenvectors of $\mathbf{S}^2 = (\mathbf{S}_1 + \mathbf{S}_2)^2$ with eigenvalue $S(S+1)$; but all states are eigenvectors of $\mathbf{S}_1^2$ and $\mathbf{S}_2^2$ with eigenvalue $3/4$, and so $S(S+1) = \mathbf{S}^2 = \mathbf{S}_1^2 + \mathbf{S}_2^2 + 2\,\mathbf{S}_1 \cdot \mathbf{S}_2 = 3/2 + 2\,\mathbf{S}_1 \cdot \mathbf{S}_2$. Replacing $\mathbf{S}_j$ by $\sigma_j/2$ gives the desired eigenvalues for $\sigma_1 \cdot \sigma_2$.

excited state of the corresponding ferromagnet.) By writing the Hamiltonian in terms of $\sigma_1 \cdot \sigma_2$, we have reformulated a Coulomb exchange interaction in the much simpler guise of a spin-spin interaction.

In a magnetically ordered solid with no external magnetic field, we sum the above spin Hamiltonian $\mathcal{H}_{\text{spin}} = \frac{J}{2}\sigma_1 \cdot \sigma_2$ over all pairs of electrons to obtain the Heisenberg (or Heisenberg-Dirac, or Heisenberg-Dirac-Van Vleck) Hamiltonian:

$$\mathcal{H} = -\frac{1}{2}\sum_{i,j} J_{ij}\ \sigma_i \cdot \sigma_j.$$

For a solid with localized electrons, the exchange integral J_{ij} falls off rapidly with increasing distance between the two electrons, and is the same for all nearest neighbor pairs of electrons, so that we may approximate the Heisenberg Hamiltonian by $\mathcal{H} = -\frac{J}{2}\sum_{\text{n.n.}} \sigma_i \cdot \sigma_j$, where the sum is over all nearest neighbor pairs. Although we have derived this Hamiltonian using electrons, it may be used more generally for particles with arbitrary spin, not just for spin-1/2 fermions. It is in the spin-1/2 case, however, that a direct connection to random walks and transition matrices may be made, as we will see in Section 4.1.

Often, due to the anisotropic nature of atoms and atom arrangements in the magnetically ordered solid, it is not equally easy to achieve magnetization in all directions. In this case, we use a generalized, anisotropic Heisenberg Hamiltonian:

$$\mathcal{H} = -\frac{1}{2}\sum_{\text{n.n.}}(J_x\sigma_i^x\sigma_j^x + J_y\sigma_i^y\sigma_j^y + J_z\sigma_i^z\sigma_j^z). \tag{2.4}$$

This is the general Heisenberg XYZ Hamiltonian. When there is one preferred direction, we align this direction with the z axis; then $J_x = J_y$, and we obtain the Heisenberg XXZ model. As a side note, if we set $J_x = J_y = 0$ in the XXZ model, we obtain the Ising model, a favorite among statistical physicists and combinatorialists alike for its understandability, relative simplicity, and applicability.

We will, however, only use the isotropic version of the Hamiltonian, the XXX model, in which $J_x = J_y = J_z$. As indicated in Section 2.2, we will further restrict our attention to the one-dimensional case with periodic boundary conditions. In this scenario, there are N electrons arranged on a line with periodic boundary conditions (or, alternatively, in a circular ring), so that each electron has two nearest neighbors. Bethe's 1931 paper [2] addresses precisely this case.

2.4 Physical solution of the Heisenberg model: magnons

From now on, we will only be working with the one-dimensional Heisenberg XXX model on N spins, so that, in particular, $J_x = J_y = J_z = J$. In addition, assume that $J > 0$, the ferromagnetic model; then we may set $J = 1$ without loss of generality:

$$\mathcal{H} = -\frac{1}{2}\sum_{j=1}^{N}(\sigma_j^x\sigma_{j+1}^x + \sigma_j^y\sigma_{j+1}^y + \sigma_j^z\sigma_{j+1}^z). \tag{2.5}$$

To find the ground state (lowest energy eigenvalue) and low-lying excitations (energies above but close to the ground state) of the ferromagnetic XXX model, solid state physics introduces the notion of a spin wave, or magnon. Although the analysis can be done in two or three dimensions, we will restrict our attention, as before, to a one-dimensional chain with periodic boundary conditions. Some of this section is based on the analysis of magnons in Keffer [11] and Mattis [13].

In the terminology of the previous section, it is an interesting and well-studied problem of physics to compute, or estimate, the energies and eigenvectors corresponding to the ground state and "elementary" excitations of the Hamiltonian $\mathcal{H}$ defined by (2.5). The ground state corresponds to the eigenvector and eigenvalue of $\mathcal{H}_0$, while elementary excitations correspond to the diagonalization of $\mathcal{H}_1$ and are known as spin waves or, synonymously, magnons. (Strictly speaking, magnons are quantized spin waves, but this is an issue of semantics.) Following a hypothesis of Felix Bloch, physicists then proceed to estimate eigenvalues and eigenvectors for $\mathcal{H}_2$ by essentially combining two spin waves to produce "two-magnon" states; these, along with a small number of "bound" states, give approximate eigenvalues and eigenvectors for $\mathcal{H}_2$. For values of r beyond 2, more spin waves can be combined to give an approximate diagonalization of $\mathcal{H}_r$, but with decreasing accuracy. In Section 3, we will approach the diagonalization of $\mathcal{H}_r$ from a rigorous mathematical standpoint, using the Bethe *ansatz*, but only for $r = 1$ and $r = 2$.

We first discuss the ground state of $\mathcal{H}$. The vector $|-\rangle$, representing a state with all spins up, is clearly an eigenstate of $\mathcal{H}$, since it alone generates $\mathfrak{H}_0$; an easy computation reveals that the corresponding energy eigenvalue is $-N/2$. This energy level is in fact $(N+1)$–fold degenerate: if $0 \le r \le N$, then $\sum_{1 \le x_1 < \cdots < x_r \le N} |x_1 \cdots x_r\rangle$ is also an eigenvector of $\mathcal{H}$ with eigenvalue $-N/2$. It is not hard to show (see, e.g., Pathria [14, p. 703]) that all eigenvalues of $\mathcal{H}$ are at least $-N/2$, so that this is in fact the ground state energy. (This statement is clear when translated into the language of Markov chains—see Section 4.1—in which it merely says that the largest eigenvalue of the transition matrix of the interchange process is 1.) Accordingly, we will write $E_0 = -N/2$ in this section and subsequent ones.

At or near absolute zero, our one-dimensional chain of spins will be in its ground state, which may be forced to be $|-\rangle$ (all spins aligned) if we apply an infinitesimal external magnetic field to break the ground state degeneracy. If we increase the temperature, the chain may be excited out of the ground state and into one of its low-lying excitations. The most elementary excitations involve states with one spin reversed, $|x_j\rangle$. None of these pure states are eigenvectors of $\mathcal{H}$, but various linear combinations of the $|x_j\rangle$ are. Indeed, the formulation of $\mathcal{H}$ from (2.3) readily gives the following criterion: $\psi = \sum_{j=1}^{N} a(j)|x_j\rangle$ is an eigenvector of $\mathcal{H}$ with eigenvalue E if and only if

$$Ea(j) = (2 - N/2)a(j) - a(j-1) - a(j+1)$$

for all j. (Here we impose the periodic boundary condition $a(j+N) = a(j)$.) This is a simple difference equation with solution $a(j) = Ae^{ipj}$, where p is determined by the boundary condition, which gives $p = 2k\pi/N$, $k = 0, \ldots, N-1$. Thus we have the N normalized

eigenvectors

$$\psi_p = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} e^{ipj} |x_j\rangle, \tag{2.6}$$

which give corresponding energy eigenvalues

$$E = E_0 + 2(1 - \cos p). \tag{2.7}$$

These eigenvectors are called spin waves, or magnons. (The former name arises from their semiclassical interpretation as chains of precessing spin vectors—see, e.g., Ashcroft and Mermin [1]; the latter name is perhaps by analogy with phonons, which come from elastic vibrations in crystals, and are quite similar in behavior to magnons.) The energy required to excite the system from its ground state to a spin wave is given by $2(1 - \cos p)$, which we will abbreviate in this section by $E(p)$. It is easy to check that these spin waves form an orthonormal basis of $\mathcal{H}_1$.

To those with some background in quantum mechanics, the formula (2.6) for a spin wave should look familiar, since it is nearly identical to the formula for the wave function ψ of a free particle in one dimension with periodic boundary condition $\psi(x + N) = \psi(x)$: $\psi(x) = e^{ipx}/\sqrt{N}$, where p is as above. In this case, p is the momentum, or wave number (since we have set $\hbar = 1$, these are the same), of the particle; in other words, ψ is an eigenfunction of the momentum operator $\mathbf{p} = -i(d/dx)$ with eigenvalue p. By analogy, we now define a "discrete" momentum operator acting on $\mathfrak{H}$, so that we can apply the familiar result from quantum mechanics that momentum commutes with a translation-invariant Hamiltonian.

Consider the site translation operator T on $\mathfrak{H}$ defined by

$$T|x_1 x_2 \cdots x_r\rangle = |x_1 + 1, x_2 + 1, \ldots, x_r + 1\rangle,$$

where the right hand side denotes $|1, x_1 + 1, x_2 + 1, \ldots, x_{r-1} + 1\rangle$ if $x_r = N$. Any eigenvector ψ of T will satisfy $T\psi = e^{ip}\psi$ for some p; here p is real, since T is unitary, and unique up to integer multiples of 2π, and is known as the (total) momentum of ψ. Now $\mathcal{H}$ is translation invariant, which means that T commutes with $\mathcal{H}$, and so momentum is a good quantum number. Thus we can find a basis of $\mathfrak{H}$ consisting of simultaneous eigenvectors of $\mathcal{H}$ and T; in particular, we can find a basis of $\mathfrak{H}_1$ of eigenvectors of $\mathcal{H}_1$ and T. Indeed, ψ_p is an eigenvector of T with eigenvalue e^{ip}, so that it has momentum p according to our definition of momentum. Under this interpretation, (2.7) becomes a so-called dispersion relation between energy E and momentum p.

Thus the "elementary" excitations of $\mathcal{H}$ are given by spin waves. Perhaps it is not surprising that the eigenvectors of $\mathcal{H}_1$ have this form; after all, with the translational symmetry inherent in the model, the eigenvectors must show a similar symmetry. It is more surprising, however, that further excitations of $\mathcal{H}$—in particular, most of the eigenvectors of $\mathcal{H}_2$—are given approximately by superpositions of spin waves, as was first postulated by Felix Bloch in 1930.

Consider a vector $\psi = \sum_{x_1<x_2} a(x_1, x_2)|x_1 x_2\rangle$ in $\mathfrak{H}_2$, and suppose that this is an eigenvector of $\mathcal{H}_2$ with eigenvalue E. For ease of notation, extend the definition of $a(x_1, x_2)$

symmetrically: if $x_2 < x_1$, define $a(x_1, x_2) = a(x_2, x_1)$, and take x_1, x_2 modulo N. Application of $\mathcal{H}_2$ to ψ leads to a series of equations that the $a(x_1, x_2)$ must satisfy:

$$\begin{aligned} &(E - E_0 - 4)a(x_1, x_2) + a(x_1, x_2 - 1) + a(x_1, x_2 + 1) + a(x_1 + 1, x_2) + a(x_1 - 1, x_2) \\ = &\begin{cases} 0 & \text{if } x_2 \neq x_1 \pm 1 \\ a(x_1, x_1) + a(x_2, x_2) - a(x_1, x_2) - a(x_2, x_1) & \text{if } x_2 = x_1 \pm 1 \end{cases} \end{aligned} \tag{2.8}$$

whenever $x_1 \neq x_2$. Here the meaningless $a(x_1, x_1)$ and $a(x_2, x_2)$ cancel on both sides of the equation, whenever they appear, unless $x_1 = x_2$. The analysis becomes simpler if we postulate that (2.8) holds for all x_1 and x_2 by imposing periodic boundary conditions (that is, $a(x_1, x_2)$ is unchanged when x_1 or x_2 changes by a multiple of N), and letting (2.8) when $x_1 = x_2$ define the "unphysical" quantities $a(x_1, x_1)$.

In Section 3.3, we will solve (2.8) exactly, using the Bethe *ansatz*. Here, we solve it approximately in order to gain insight about the solutions. If the right hand side of (2.8) were zero for all (x_1, x_2) rather than for most pairs, it would be solvable by the "plane wave" ψ_{p_1,p_2} given by

$$a(x_1, x_2) = e^{i(p_1 x_1 + p_2 x_2)} + e^{i(p_2 x_1 + p_1 x_2)}, \tag{2.9}$$

where p_1 and p_2 are given by periodic boundary conditions, which postulate that $Np_j/2\pi \in \mathbb{Z}$. (Since we have required $a(x_1, x_2) = a(x_2, x_1)$, we cannot have unequal constant factors multiplying the exponentials, as we will have in (3.8) in Section 3.3.) Intuitively, ψ_{p_1,p_2} is the symmetrized combination (in this case, a product) of two spin waves $\psi_{p_1}(x_1)$ and $\psi_{p_2}(x_2)$, which have momenta p_1 and p_2, respectively. This would yield an energy eigenvalue E, from (2.8), satisfying

$$E - E_0 = 2(1 - \cos p_1) + 2(1 - \cos p_2) = E(p_1) + E(p_2), \tag{2.10}$$

which may be interpreted (cf. (2.7)) as the combined energy of two superimposed spin waves with momenta p_1 and p_2. (Recall that $E(p)$ is the energy associated with a spin wave of momentum p.)

As with $\mathcal{H}_1$, we may take advantage of the translation invariance of $\mathcal{H}_2$ to define a "total" momentum operator (so called to distinguish it from the pseudo-momenta p_1 and p_2). Since $T\psi_{p_1,p_2} = e^{i(p_1+p_2)}\psi_{p_1,p_2}$, the vector ψ_{p_1,p_2} has total momentum $P = p_1 + p_2$.

Some of the ψ_{p_1,p_2} are in fact eigenvectors of $\mathcal{H}_2$. When $p_2 = 0$ (or, similarly, when $p_1 = 0$), the vector $\psi_{p_1,0}$ given by (2.9) is an exact eigenvector of $\mathcal{H}_2$; if $\sigma^-_{\text{tot}} = \sum_j \sigma^-_j$ represents the total spin lowering operator, then $\psi_{p_1,0} = \sqrt{N}\sigma^-_{\text{tot}}\psi_{p_1}$, where ψ_{p_1} is an eigenvector of $\mathcal{H}_1$, as defined above. (Note that $\psi_{p_1,0}$ is an eigenvector of $\mathcal{H}_2$ because σ^-_{tot} commutes with $\mathcal{H}$.) In general, applying σ^-_{tot} to an eigenvector of $\mathcal{H}_r$ yields an eigenvector of $\mathcal{H}_{r+1}$; see Section 6.2.

When $p_1, p_2 \neq 0$, the vectors ψ_{p_1,p_2} are not eigenvectors of $\mathcal{H}$, but we may try to combine vectors with the same total momentum ($P = p_1 + p_2$ constant) to obtain eigenvectors of $\mathcal{H}$ with well-defined total momentum. Change coordinates by writing $p = (p_1 - p_2)/2$, $x = x_1 - x_2$, $X = (x_1 + x_2)/2$, so that $a(x_1, x_2) = e^{iPX}(e^{ipx} + e^{-ipx})$. (These new coordinates have a physical interpretation: X is the center-of-mass coordinate, x is the relative coordinate, and p is a sort of relative momentum.) Now form a linear combination ψ over states with

constant P and varying p, where p attains all possible values:

$$\psi(x_1, x_2) = e^{iPX}(\sum_p e^{ipx} f(p)) = e^{iPX} F(x),$$

where $f(p) = f(-p)$ and this equation defines $F(x)$ (so that $F(x) = F(-x)$). Substituting into the characteristic equation (2.8), multiplying by e^{-ipx}, and summing over x gives

$$(E - E_0 - E(p_1) - E(p_2))f(p) = \frac{4}{N} \sum_{p'} (\cos x) \left(\cos \frac{P}{2} - \cos p' \right) f(p'). \tag{2.11}$$

This series of equations may be further analyzed (see Mattis [13]) to find $f(p)$ and a more exact value for E. For our purposes, it suffices to note that (2.11) implies that, up to order $1/N$, $E \approx E_0 + E(p_1) + E(p_2)$, in accordance with (2.10).

When p_1 and p_2 are real, we have each momentum independently ranging more or less uniformly in $[0, 2\pi)$ (recall that originally $Np_j/2\pi \in \mathbb{Z}$). We then have a "continuum" of so-called two-magnon states with energy eigenvalues given approximately by (2.10). Since we may write $E(p_1)+E(p_2) = 4(1-\cos(P/2)\cos p)$, the energies of two-magnon states range, for fixed P, between the extremes $E_0 + 4(1 - \cos(P/2))$ and $E_0 + 4(1 + \cos(P/2))$.

There are also states in which p_1 and p_2 are not real. (Note that total momentum $P = p_1 + p_2$ is real, even if p_1 and p_2 are not.) These are known as "bound states," and for given P, their energies lie outside (in fact, below) the corresponding range of two-magnon energies.

We will not go into further detail here, since we will precisely calculate the eigenvectors and eigenvalues of $\mathcal{H}_2$ in Section 3, using the Bethe *ansatz*. In Section 4.4, we will rederive, in a mathematically rigorous fashion, our analysis from this section regarding two-magnon and bound states.

3 Diagonalization via the Bethe *ansatz*

Having discussed physical aspects of the Heisenberg model, we now proceed to effect a mathematical diagonalization of the one-dimensional XXX model by using the Bethe *ansatz*. Section 3.1 introduces the formulation of the Bethe *ansatz* relevant to the XXX model, and Sections 3.2 and 3.3 carry out the diagonalizations of $\mathcal{H}_r$ in the case of one ($r = 1$) and two ($r = 2$) opposite spins, respectively.

3.1 The mathematical Bethe *ansatz*

The exact solution of the Heisenberg isotropic ferromagnetic model hinges on the powerful *ansatz*, or hypothesis, first postulated by Bethe [2] in 1931. Originally used to hypothesize the form of the eigenvectors of the XXX model, the Bethe *ansatz* has been generalized successfully to give the form of eigenvectors in many other models, including delta-function interaction potentials, analyzed by C. N. Yang and C. P. Yang, and ice-type models, analyzed by Lieb and Baxter. C. N. Yang [17] gives three equivalent intuitive physical descriptions of the Bethe *ansatz*, in its greatest generality: the wave function of a system is a finite sum of pure exponentials in each of a number of regions into which the coordinate space of the system is divided; the system in question displays reflection but no diffraction; and the system obeys a "large number of conservation laws."[2] Of these descriptions, the first comes closest to the version of the Bethe *ansatz* which we will use, which roughly states that the wave function in the region $\{x_1 < \cdots < x_N\}$ is a superposition of exponentials.

We now describe the "classical" Bethe *ansatz*, as applied to the XXX model. Note that the formulation below of the Bethe *ansatz* does not purport to be a formula for eigenvectors; it is merely a tool which makes the task of finding eigenvectors easier by suggesting that they take a certain form. Using the Bethe *ansatz*, we may easily find a number of eigenvectors, which we must then count to verify that they constitute all eigenvectors.

Bethe *ansatz* 3.1 *The eigenvectors*

$$\psi = \sum_{1 \le x_1 < \cdots < x_r \le N} a(x_1, \ldots, x_r) |x_1 \cdots x_r\rangle$$

of $\mathcal{H}_r$ have coefficients of the form

$$a(x_1, \ldots, x_r) = \sum_{\pi \in S_r} A_\pi \exp\left[i \sum_{j=1}^{r} p_{\pi(j)} x_j \right],$$

where $p_1, \ldots, p_r$ are r unequal numbers, S_r is the symmetric group on r elements, and the A_π are complex numbers depending on π.

It is straightforward to show that ψ from the Bethe *ansatz* has a well-defined total momentum (as defined in Section 2.4) of $p_1 + \cdots + p_m$.

[2] Yang illustrates the importance of the Bethe *ansatz* by using it in [17] to deduce the central Yang-Baxter equation in statistical mechanics.

From the Bethe *ansatz*, it is possible to deduce heuristically a set of conditions which determine the energy eigenvalues of $\mathcal{H}_m$; see Izyumov and Skryabin [9] for this nonrigorous argument, as well as a more general result for the XXZ model. Here we will simply quote the result from [9]: the eigenvalues of $\mathcal{H}_r$ are given by

$$E = E_0 + \sum_{j=1}^{m} \frac{1}{\lambda_j^2 + 1/4}, \tag{3.1}$$

where the λ_j satisfy the series of equations

$$\left(\frac{\lambda_j + i/2}{\lambda_j - i/2}\right)^N = \prod_{1 \le \ell \le N,\ \ell \ne j} \frac{\lambda_j - \lambda_\ell + i}{\lambda_j - \lambda_\ell - i} \tag{3.2}$$

for each $j = 1, \dots, N$; λ_j here is related to p_j from the Bethe *ansatz* via the relation

$$p_j = 2\cot^{-1}(2\lambda_j) = \frac{1}{i}\log\frac{\lambda_j + i/2}{\lambda_j - i/2}. \tag{3.3}$$

Although we will use equations (3.1), (3.2), and (3.3) in Sections 3.2 and 3.3 to find precise eigenvectors of $\mathcal{H}_r$, we will independently verify that the vectors thus obtained are in fact eigenvectors, thus showing that the equations are correct in these cases. These equations, like the Bethe *ansatz* itself, are simply a convenience which allow us to find eigenvectors of $\mathcal{H}_r$ more easily than otherwise.

3.2 Diagonalization of $\mathcal{H}_1$

As a first application of the Bethe *ansatz*, we will diagonalize the Heisenberg XXX Hamiltonian in the case of one opposite spin, i.e., when $r = 1$. This relatively simple example is easy to approach rigorously; we follow the approach of Izyumov and Skryabin [9]. In Section 3.3, we treat the case $r = 2$, which is more unwieldy and whose mathematical formalism has been inadequately addressed in the literature. The remaining cases $3 \le r \le N/2$ are too complicated to be precisely treated in this paper. Indeed, the exact diagonalization of K_r is equivalent to the solution of an r-body problem, which is essentially impossible when $r \ge 3$: "the analysis of 3- or more-magnon states requires the solution of a full many-body problem" [13, p. 138].

We consider the Bethe *ansatz* in the case $r = 1$, and then verify that it is correct in this case. Equation (3.2), which we stated without proof, simply becomes $\left(\frac{\lambda+i/2}{\lambda-i/2}\right)^N = 1$; if p is defined as in (3.3), then $e^{ipN} = 1$, and so $p = 2k\pi/N$ for some $k = 0, \dots, N-1$. The Bethe *ansatz* then postulates that the eigenvectors ψ of $\mathcal{H}_1$ are given by

$$\psi = A\sum_{x=1}^{N} e^{ipx}|x\rangle.$$

It is easy to check that when $p = 2k\pi/N$, the vector ψ is indeed an eigenfunction of $\mathcal{H}_1$, with eigenvector $E = E_0 + 2(1 - \cos p)$; we did this, more or less, in Section 2.4. (Recall that E_0

is defined by $E_0 = -N/2$.) The N eigenvectors thus obtained form an orthogonal basis of $\mathfrak{H}_1$ (orthonormal if $A = \sqrt{N}$) with respect to the usual complex scalar product. This proves the following result, already stated in Section 2.4.

Proposition 3.2 *The N eigenvalues of $\mathcal{H}_1$ are given by*

$$E = E_0 + 2(1 - \cos(2k\pi/N)),$$

where $k = 0, \ldots, N-1$.

3.3 Diagonalization of $\mathcal{H}_2$

In the mathematical physics literature, the use of the Bethe *ansatz* to diagonalize the Heisenberg XXX Hamiltonian is typically presented without much rigor; see, for instance, Bethe's original paper [2] or Izyumov and Skryabin's detailed exposition of the Bethe *ansatz* in [9]. The problem of showing that the eigenvectors produced by the Bethe *ansatz* give a complete diagonalization of the Heisenberg Hamiltonian is often treated lightly or not at all. In this section, we will rigorously treat the diagonalization of $\mathcal{H}_r$ in the case of two opposite spins, i.e., when $r = 2$.

In the case $r = 2$, the series of equations (3.2) becomes

$$\left(\frac{\lambda_1 + i/2}{\lambda_1 - i/2}\right)^N = \frac{\lambda_1 - \lambda_2 + i}{\lambda_1 - \lambda_2 - i} = \left(\frac{\lambda_2 - i/2}{\lambda_2 + i/2}\right)^N. \tag{3.4}$$

This equation and the relation between p_j and λ_j from (3.3) imply, in particular, that $e^{iNp_1p_2} = 1$, so that $e^{ip_1p_2} = \omega^2$, where $\omega = e^{i\pi k/N}$ for some $k = 0, \ldots, N-1$. If we then set $x = e^{ip_2}/\omega = \omega/e^{ip_1}$, and throw out the nonsensical root $x = 0$, then (3.4) becomes an N-th degree polynomial equation in x:

$$(\omega + \omega^{-1})x^N - 2x^{N-1} - 2\omega^N x + (\omega^{N+1} + \omega^{N-1}) = 0.$$

In terms of x, the expression (3.1) for energy becomes

$$E = E_0 + 4 - (\omega + \omega^{-1})(x + x^{-1}).$$

The following result verifies the Bethe *ansatz* predictions of (3.1) and (3.2) by showing that the above formulas give essentially all of the eigenvalues of $\mathcal{H}_2$.

Proposition 3.3 *The $\binom{N}{2}$ eigenvalues of $\mathcal{H}_2$ are given by*

$$E(k, x) = E_0 + 4 - 2\left(\cos\frac{k\pi}{N}\right)\left(x + \frac{1}{x}\right), \tag{3.5}$$

where $x \neq 0, \pm 1$ is a root of the polynomial equation

$$q_k(x) \equiv \left(\cos\frac{k\pi}{N}\right)x^N - x^{N-1} - (-1)^k x + (-1)^k\left(\cos\frac{k\pi}{N}\right) = 0 \tag{3.6}$$

and k runs from 0 to $N-1$, along with the extra eigenvalues $E = E_0$ for all N, and $E = E_0 + 2$ when N is even. (With respect to multiplicity, only one of each pair of $E(k, x)$ and $E(k, 1/x)$ is counted as an eigenvalue.)

Here (3.6) defines the polynomial $q_k(x)$. In order to prove the proposition, we first establish a lemma essentially due to Bethe [2].

Lemma 3.4 *Given $k = 0, \ldots, N-1$ and x satisfying $q_k(x) = 0$, there exists an eigenvector $\psi_{k,x}$ of $\mathcal{H}_2$ with eigenvalue $E(k,x)$ defined by (3.5).*

Proof. Our strategy is to use the Bethe *ansatz* as motivation to find a potential eigenvector of $\mathcal{H}_2$, and then to derive $q_k(x) = 0$ as a necessary and sufficient condition for our vector to be an eigenvector with the desired eigenvalue. The Bethe *ansatz* hypothesizes that eigenvectors of $\mathcal{H}_2$ will have the form

$$\psi = \sum_{1 \le x_1 < x_2 \le N} a(x_1, x_2) |x_1 x_2\rangle, \tag{3.7}$$

where

$$a(x_1, x_2) = A_{12} e^{i(p_1 x_1 + p_2 x_2)} + A_{21} e^{i(p_2 x_1 + p_1 x_2)} \tag{3.8}$$

and $p_1 \neq p_2$. For ease of notation, introduce new variables P_1 and P_2 where $P_j = e^{ip_j}$, so that (3.8) becomes

$$a(x_1, x_2) = A_{12} P_1^{x_1} P_2^{x_2} + A_{21} P_2^{x_1} P_1^{x_2}. \tag{3.9}$$

We now rewrite (2.8), the conditions on the coefficients $a(x_1, x_2)$ for ψ to be an eigenfunction of $\mathcal{H}_2$ with eigenvalue E:

$$Ea(x_1, x_2) = (4 - N/2)\, a(x_1, x_2) - a(x_1 - 1, x_2) - a(x_1 + 1, x_2) - a(x_1, x_2 - 1) - a(x_1, x_2 + 1) \tag{3.10}$$

if $x_2 \neq x_1 + 1$, and

$$Ea(x_1, x_1 + 1) = (2 - N/2)\, a(x_1, x_1 + 1) - a(x_1 - 1, x_1 + 1) - a(x_1, x_1 + 2). \tag{3.11}$$

The function $a(x_1, x_2)$ is only defined when x_1 and x_2 are integers with $1 \le x_1 < x_2 \le N$. Equation (3.9) makes sense for all x_1 and x_2, however, so extend the domain of $a(x_1, x_2)$ to all integers $x_1 < x_2$; we will later impose the necessary boundary condition, that $a(x_2, x_1 + N) = a(x_1, x_2)$. (Since (3.9) is not necessarily symmetric in x_1 and x_2, we cannot extend the domain of $a(x_1, x_2)$ by symmetry to all x_1 and x_2, as we did in Section 2.4.)

It is straightforward to check that (3.9) satisfies (3.10) for arbitrary A_{12} and A_{21}, when

$$E = E_0 + 2(1 - \cos p_1) + 2(1 - \cos p_2). \tag{3.12}$$

Thus (3.10) is satisfied for all x_1 and x_2 (not necessarily with $x_2 \neq x_1 + 1$) when (3.12) holds. A comparison of (3.11) with the case $x_2 = x_1 + 1$ in (3.10) yields the additional condition

$$2a(x_1, x_1 + 1) = a(x_1, x_1) + a(x_1 + 1, x_1 + 1), \tag{3.13}$$

which is then equivalent to (3.11). On substituting (3.9), we find that (3.13), in turn, is equivalent to the relation

$$\frac{A_{12}}{A_{21}} = -\frac{2P_1 - 1 - P_1 P_2}{2P_2 - 1 - P_1 P_2}. \tag{3.14}$$

Finally, the boundary condition $a(x_1, x_2) = a(x_2, x_1 + N)$ holds, from the definition (3.9), exactly when

$$\frac{A_{12}}{A_{21}} = P_1^N = P_2^{-N}. \tag{3.15}$$

We may eliminate the auxiliary variables A_{12} and A_{21} from (3.14) and (3.15) to get

$$P_1^N = P_2^{-N} = -\frac{2P_1 - 1 - P_1P_2}{2P_2 - 1 - P_1P_2}. \tag{3.16}$$

We conclude that if P_1, P_2, and E satisfy (3.12) and (3.16), then the vector with coefficients defined by (3.9) is an eigenvector of $\mathcal{H}_2$ with eigenvalue E. Now given k and $x \neq \pm 1$ satisfying (3.6), we may define

$$P_1 = \omega/x, \quad P_2 = \omega x, \tag{3.17}$$

where $\omega = e^{i\pi k/N}$. (The requirement that $x \neq \pm 1$ ensures that $P_1 \neq P_2$, so that $a(x_1, x_2)$ is not identically zero. As a side note, (3.17) implies that the total momentum $p_1 + p_2$, as defined in Section 2.4, is equal to $2k\pi/N$, a fact we will use in the proof of Proposition 3.3.) With this definition, (3.16) becomes precisely the given condition (3.6), while (3.12) becomes (3.5). Thus our values for k and x produce an eigenvector $\psi_{k,x}$ defined by (3.7) and (3.9) with eigenvalue $E(k, x)$, as desired. □

Proof of Proposition 3.3. We first count the number of eigenvalues and eigenvectors we obtain through the construction of Lemma 3.4. The polynomial equation (3.6), $q_k(x) = 0$, has N roots unless N is even and $k = N/2$, in which case it has $N - 2$ nonzero roots. Note that if x is a root of q_k, then $1/x$ does also, and that $\psi_{k,x} = \psi_{k,1/x}$ in this case. (The transformation $x \to 1/x$ merely interchanges P_1 and P_2.) We may check that q_k has no double roots when $k \neq 0$ by simultaneously setting $q_k = dq_k/dx = 0$; we obtain a quadratic in x,

$$(N-1)\left(\cos\frac{k\pi}{N}\right)x^2 - N\left(\sin\frac{k\pi}{N}\right)^2 x + \left(1 - N\cos\frac{k\pi}{N}\right) = 0,$$

whose two roots can be verified not to satisfy $q_k(x) = 0$ when $k \neq 0$. Also note that $x = 1$ is a root of q_k exactly when $k = 0$ (double root) or k is odd, while $x = -1$ is a root exactly when $N + k$ is odd.

Suppose that N is even. From the above observations, a simple counting argument yields that a given value of $k = 0, \ldots, N-1$ will produce $N/2 - 1$ eigenvectors if $k = 0$, $k = N/2$, or k is odd, and $N/2$ eigenvectors if k is even and $k \neq 0, N/2$, for a total of $\left(\frac{N}{2} - 1\right)\left(\frac{N}{2} + 2\right) + \left(\frac{N}{2}\right)\left(\frac{N}{2} - 2\right) = \binom{N}{2} - 2$ eigenvectors. Two additional eigenvectors are given by ψ_0, defined by $a(x_1, x_2) \equiv 1$, with eigenvalue $E_0 = -N/2$, and ψ_{even}, defined by

$$a(x_1, x_2) = \begin{cases} (-1)^{x_1} & \text{if } x_2 = x_1 + 1 \\ 1 & \text{if } x_1 = 1, x_2 = N \\ 0 & \text{otherwise,} \end{cases}$$

with eigenvalue $E_0 + 2$. Note that we have chosen ψ_0, the constant eigenvector, so that we may regard it as an eigenvector of the form $\psi_{k,x}/2$, where $k = 0$ and $x = 1$.

Similarly, if N is odd, we find that a given value of k will produce $(N-1)/2$ eigenvectors when $k \neq 0$, and $(N-3)/2$ eigenvectors when $k = 0$, for a total of $\binom{N}{2} - 1$ eigenvectors. One additional eigenvector is given, as above, by ψ_0, defined by $a(x_1, x_2) \equiv 1$, with eigenvalue E_0.

Thus, for either parity of N, we obtain $\binom{N}{2}$ eigenvectors. To complete the diagonalization of $\mathcal{H}_2$, it suffices to show that these eigenvectors are linearly independent. In fact, all pairs of eigenvectors are orthogonal under the usual complex scalar product $(\cdot, \cdot)$, as we now show. Recall the site translation operator T from Section 2.4, which is defined on $\mathfrak{H}_2$ by $T|x_1 x_2\rangle = |x_1 + 1, x_2 + 1\rangle$. Any eigenvector $\psi_{k,x}$ of $\mathcal{H}_2$ produced by the Bethe *ansatz* and Lemma 3.4 is, by virtue of its form (3.8), an eigenvector of T as well, with eigenvalue $e^{p_1+p_2} = e^{2i\pi k/N}$. (In the language of Section 2.4, $\psi_{k,x}$ has total momentum $2\pi k/N$.) Now consider two eigenvectors $\psi_{k,x}$ and $\psi_{\tilde{k},\tilde{x}}$ produced by Lemma 3.4. If $k \neq k'$, then $\psi_{k,x}$ and $\psi_{\tilde{k},\tilde{x}}$ are eigenvectors of T with different eigenvalues; since T is hermitian, the two vectors are orthogonal. If $k = k'$, on the other hand, then we claim that $\psi_{k,x}$ and $\psi_{\tilde{k},\tilde{x}}$ are eigenvectors of $\mathcal{H}$ with different eigenvalues, so that they are again orthogonal. Indeed, their eigenvalues with respect to $\mathcal{H}$ are the same, by (3.5), only if $x + 1/x = \tilde{x} + 1/\tilde{x}$, which happens only if $\tilde{x} = x$ or $\tilde{x} = 1/x$, a case we have already ruled out by noting that q_k has no double roots, and allowing only one of each pair of $\psi_{k,x}$ and $\psi_{k,1/x}$. Thus all pairs $\psi_{k,x}$ and $\psi_{\tilde{k},\tilde{x}}$ are orthogonal.

The "extra" eigenvectors ψ_0 and ψ_{even} are orthogonal to all of these eigenvectors. If we write ψ_0 in the form $\psi_{k,x}/2$ where $k = 0$, $x = 1$, then the above argument shows that ψ_0 is orthogonal to all of the $\psi_{k,x}$. In addition, if N is even, then

$$(\psi_{k,x}, \psi_{\text{even}}) = \sum_{x_1=1}^{N} (-1)^{x_1} a(x_1, x_1 + 1) = \sum_{x_1=1}^{N} (-1)^{x_1} \left(\frac{a(x_1, x_1) + a(x_1 + 1, x_1 + 1)}{2} \right) = 0,$$

where we have used (3.13) and the periodic boundary condition for $a(x_1, x_2)$.

Thus all eigenvectors are pairwise orthogonal, and hence the entire collection of eigenvectors is linearly independent, which completes the diagonalization and the proof of Proposition 3.3. $\square$

4 Applications of the diagonalization

This section is devoted to consequences of the diagonalization of $\mathcal{H}_1$ and $\mathcal{H}_2$ effected in Section 3. In Section 4.1, we show that the Heisenberg Hamiltonian is directly related to the transition matrix of the interchange process, so that our diagonalization from Section 3 may be applied to the interchange process. Section 4.2 gives some preliminary results necessary for Sections 4.3 and 4.4; Section 4.3 calculates the so-called spectral gap of the interchange chain, which essentially determines how quickly the chain becomes random, and Section 4.4 rederives results given by Izyumov and Skryabin [9] on the asymptotic distribution of eigenvalues of the interchange model when the number of cards is large.

4.1 The interchange model K_r

Consider the $\binom{N}{r}$-dimensional complex vector space generated by the $\binom{N}{r}$ configurations of r red and $N-r$ black cards in a circle. Through the identification of a red card in position j with the down spin vector e_j^- and a black card in position j with e_j^+, this vector space is canonically isomorphic to the space $\mathfrak{H}_r$ defined in Section 2.2.

Now consider the following random walk on $\mathfrak{H}_r$, which we call the interchange process on $\mathbb{Z}_N$. At each step, choose a pair of neighboring cards (sites) uniformly at random, and switch the cards (spins occupying those sites). This defines a reversible, irreducible Markov chain with uniform stationary distribution $\mu \equiv 1/\binom{N}{r}$. (For the precise definitions of these terms, see Section 5.1.) Let K_r be the associated $\binom{N}{r} \times \binom{N}{r}$ transition matrix.

We may define K_r formally as follows. Identify a basis element of $\mathfrak{H}_r$ by the r-element subset of $\mathbb{Z}_N$ corresponding to the positions of the red cards (down spins, e_j^-); in other words, identify $|x_1 \cdots x_r\rangle$ with $\{x_1, \ldots, x_r\} \subset \mathbb{Z}_N$. If A_1 and A_2 are two r-subsets of $\mathbb{Z}_N$, then the transition probability from A_1 to A_2 is

$$K_r(A_1, A_2) = \begin{cases} 0 & \text{if } |A_1 \cap A_2| \le r-2 \\ 0 & \text{if } |A_1 \cap A_2| = r-1 \text{ and } A_1 - A_2 = \{x_1\}, \\ & \quad A_2 - A_1 = \{x_2\}, \text{ with } x_2 - x_1 \neq \pm 1 \\ 1/N & \text{if } |A_1 \cap A_2| = r-1 \text{ and } A_1 - A_2 = \{x_1\}, \\ & \quad A_2 - A_1 = \{x_2\}, \text{ with } x_2 - x_1 = \pm 1 \\ 1 - d(A_1)/N & \text{if } A_1 = A_2, \text{ where} \\ & \quad d(A_1) = |\{(x_1, x_2) : x_1 \in A_1, x_2 \notin A_1, x_2 - x_1 = \pm 1\}|. \end{cases} \tag{4.1}$$

As a side note, we may define, in a similar manner, a more general interchange process on any finite graph G, as follows. Place r red cards and $N-r$ black cards on the N vertices of G; a step in the chain consists of choosing an edge of G uniformly at random and switching the cards on the endpoints of that edge. Our particular interchange process corresponds to letting G be a cycle on N vertices, which is naturally associated with the Cayley graph of $\mathbb{Z}_N$ with generators $\{\pm 1\}$. Subsequent references to "interchange" without further qualifiers will be understood to mean interchange on $\mathbb{Z}_N$.

In order to determine the rate of convergence to the limiting stationary distribution of the interchange Markov chain, we wish to find the eigenvalues of K_r. The crucial fact in

diagonalizing K_r is that K_r is very simply related to $\mathcal{H}_r$; we may then use the results from Sections 3.2 and 3.3 to compute the spectrum of K_r.

Proposition 4.1 *If I is the identity operator, then $K_r = I/2 - \mathcal{H}_r/N$.*

PROOF. This follows directly from (2.2) and the equation $K_r = (\sum_{j=1}^{N} P_{j,j+1})/N$, where $P_{j,j+1}$ is defined, as in Section 2.2, to be the operator switching the sites labelled j and $j+1$. □

Corollary 4.2 *The eigenvalues E of $\mathcal{H}_r$ and β of K_r are in one-to-one correspondence via the relation $\beta = 1/2 - E/N$.*

Because of Corollary 4.2, we will examine the eigenvalues of $\mathcal{H}_r$ in subsequent sections, in order to obtain information about our Markov chain K_r.

Corollary 4.3 *The eigenvalues β of K_1 are given by $\beta = 1 - \frac{2}{N}(1 - \cos\left(\frac{2k\pi}{N}\right))$, where $k = 0, \ldots, N-1$.*

Corollary 4.4 *The eigenvalues β of K_2 are given by*

$$\beta(k, x) = 1 - \frac{2}{N}\left(2 - \left(\cos\frac{k\pi}{N}\right)\left(x + \frac{1}{x}\right)\right),$$

where $k = 0, \ldots, N-1$ and $x \neq 0, \pm 1$ satisfies $q_k(x) = 0$, along with the extra eigenvalues $\beta = 1$ for all N, and $\beta = 1 - 2/N$ when N is even.

The interchange model K_r on r red cards and $N - r$ black cards has a simple interpretation when $r = 1$ (or $r = N - 1$). In this case, there is only one distinguished card; at each stage, it moves left with probability $1/N$, moves right with probability $1/N$, and stays put with probability $1 - 2/N$. This chain is therefore linearly related to the simple random walk on $\mathbb{Z}_N$; the interchange transition matrix K_r is precisely $(1 - 2/N)I$ plus $2/N$ times the simple random walk transition matrix. The eigenvalues of the simple random walk on $\mathbb{Z}_N$ are well known to be $\cos(2k\pi/N)$, $k = 0, \ldots, N-1$, and so we recover Corollary 4.3.

For other values of r, the interchange process K_r cannot be so easily analyzed, and we believe that much of our analysis of K_2 is original. Unlike the state space of K_1, which is essentially the group $\mathbb{Z}_N$, the state space X_r of K_r for higher values of r is not a group, and so we cannot use the rich theory of random walks on groups (see, e.g., Diaconis [3, §3]); X_r is, however, a homogeneous space, as we now explain. Since X_r represents the collection of configurations of r red cards and $N-r$ black cards on $\{1, \ldots, N\}$, there is a natural transitive action of the symmetric group S_N on X_r which simply permutes the arrangement of cards according to some permutation π; in other words, the action of S_N on itself descends to an action of S_N on X_r. If x_0 is some configuration in X_r, then the isotropy subgroup of x_0, or the subgroup of permutations which fix x_0, is isomorphic to $S_r \times S_{N-r}$ (permutations which permute the r red cards among themselves, and the $N - r$ black cards among themselves), and X_r is thus isomorphic to $S_N/(S_r \times S_{N-r})$. The transitive action of S_N on X_r, and the resulting identification of X_r with a quotient of S_N, makes X_r a homogeneous space, as defined by Diaconis [3, §3F].

4.2 Nature of the roots of q_k

For the remainder of Section 4, we will focus on the problem of analyzing the spectrum of $\mathcal{H}_2$, although we will also give corresponding results for $\mathcal{H}_1$ for purposes of comparison. In order to understand the eigenvalues of $\mathcal{H}_2$, we first make a few remarks about the roots of the polynomials q_k. If $x \neq 0, \pm 1$ is a root of q_k, then by Proposition 3.3, there is an eigenvalue of $\mathcal{H}_2$ which is linearly related to $x + 1/x$ by (3.5). Since $\mathcal{H}_2$ is hermitian, it has all real eigenvalues, and so $x + 1/x$ must be real. But this can happen only if x is real or $|x| = 1$. It will be convenient to abbreviate the complex unit circle $\{|x| = 1\}$ by S^1; then every root of q_k is in $\mathbb{R} \cup S^1$.

The distinction between the two types of roots of q_k—those lying in $\mathbb{R}$, and those lying in S^1—corresponds to our distinction in Section 2.4 between two-magnon and bound states, respectively. Indeed, if $x \in S^1$, then P_1 and P_2, as defined in (3.17), both lie on the complex unit circle, so that the "momenta" p_1 and p_2 are real; in our physical description from Section 2.4, the resulting eigenvector is precisely a two-magnon state. By contrast, if x is real, then P_1 and P_2 do not lie on the unit circle, so that p_1 and p_2 are not real; the resulting eigenvector is a bound state.

How many real roots does q_k have? If $x \neq \pm 1$, then $q_k(x) = 0$ if and only if

$$\frac{x^{N-1} + (-1)^k x}{x^N + (-1)^k} = \cos \frac{k\pi}{N}. \tag{4.2}$$

Analysis of the behavior of the left hand side of this equation yields the following result. Note that if N is even and $k = N/2$, then q_k has no real roots not equal to 0 or ± 1.

Proposition 4.5 *If $k \neq N/2$, then q_k has exactly two real roots not equal to ± 1 if the following condition holds:*

$$\cos \frac{k\pi}{N} \in \begin{cases} (-1, 1) & \text{if } N, k \text{ even} \\ (-\frac{N-2}{N}, \frac{N-2}{N}) & \text{if } N \text{ even, } k \text{ odd} \\ (-\frac{N-2}{N}, 1) & \text{if } N \text{ odd, } k \text{ even} \\ (-1, \frac{N-2}{N}) & \text{if } N, k \text{ odd.} \end{cases}$$

Otherwise, q_k has no real roots not equal to ± 1.

From Proposition 4.5, we see that practically all of the roots of q_k lie on S^1. If $x = e^{i\theta}$ is a root of q_k, then we may divide the numerator and denominator of (4.2) by $e^{iN\theta/2}$ to obtain

$$\cos \frac{k\pi}{N} = \begin{cases} \frac{\cos(\left(\frac{N}{2} - 1\right)\theta)}{\cos\left(\frac{N}{2}\theta\right)} & \text{if } k \text{ is even} \\ \frac{\sin(\left(\frac{N}{2} - 1\right)\theta)}{\sin\left(\frac{N}{2}\theta\right)} & \text{if } k \text{ is odd.} \end{cases} \tag{4.3}$$

Straightforward analysis of the functions $\frac{\cos(\left(\frac{N}{2} - 1\right)\theta)}{\cos\left(\frac{N}{2}\theta\right)}$ and $\frac{\sin(\left(\frac{N}{2} - 1\right)\theta)}{\sin\left(\frac{N}{2}\theta\right)}$ gives us the following result, which says that the roots of q_k in S^1 are essentially evenly spaced in S^1 with respect to argument.

Proposition 4.6 *There is exactly one root of q_k in S^1 with argument in each of the following intervals:*

k even, N even: $(\frac{\pi}{N}, \frac{3\pi}{N}), (\frac{3\pi}{N}, \frac{5\pi}{N}), \ldots, (\frac{(N-3)\pi}{N}, \frac{(N-1)\pi}{N})$;
k even, N odd: $(\frac{\pi}{N}, \frac{3\pi}{N}), (\frac{3\pi}{N}, \frac{5\pi}{N}), \ldots, (\frac{(N-4)\pi}{N}, \frac{(N-2)\pi}{N})$, *and also* $(\frac{(N-2)\pi}{N}, \pi)$ *if* $\cos\frac{k\pi}{N} < -\frac{N-2}{N}$;
k odd, N even: $(\frac{2\pi}{N}, \frac{4\pi}{N}), (\frac{4\pi}{N}, \frac{6\pi}{N}), \ldots, (\frac{(N-4)\pi}{N}, \frac{(N-2)\pi}{N})$, *and also* $(0, \frac{2\pi}{N})$ *if* $\cos\frac{k\pi}{N} > \frac{N-2}{N}$, *and* $(\frac{(N-2)\pi}{N}, \pi)$ *if* $\cos\frac{k\pi}{N} < -\frac{N-2}{N}$;
k odd, N odd: $(\frac{2\pi}{N}, \frac{4\pi}{N}), (\frac{4\pi}{N}, \frac{6\pi}{N}), \ldots, (\frac{(N-3)\pi}{N}, \frac{(N-1)\pi}{N})$, *and also* $(0, \frac{2\pi}{N})$ *if* $\cos\frac{k\pi}{N} > \frac{N-2}{N}$.

Furthermore, all roots of q_k in $S^1 - \{\pm 1\}$ are given by the roots in these intervals, along with their inverses (complex conjugates).

4.3 Spectral gap of K_1 and K_2

We now calculate the second largest eigenvalue β_1 of K_r for $r = 1$ and $r = 2$. In the literature, the absolute value of the eigenvalue β_* of a Markov chain which is second largest in absolute value, after 1, is known as the spectral gap, and its value controls the rate of convergence of the chain; see Section 5.1 for more particulars. After computing β_1, we will show that the smallest eigenvalue of K_1 (respectively K_2) is at least $1 - 4/N$ (respectively $1 - 8/N$); in particular, it is nonnegative for $N \geq 8$, and so the spectral gap β_* is β_1 in these cases.

By Corollary 4.2, the two largest eigenvalues of K_r correspond to the ground state and lowest excitation of the chain of fermions described by $\mathcal{H}_r$. The result for $r = 1$ follows immediately from Proposition 3.2, while the result for $r = 2$, as usual, requires more work.

Proposition 4.7 *The ground state and lowest excitation energy eigenvalues of $\mathcal{H}_1$ are given by $E_0 = -N/2$, with multiplicity 1, and*

$$E_1 = E_0 + 4\left(\sin\frac{\pi}{N}\right)^2,$$

with multiplicity 2.

Proposition 4.8 *The ground state and lowest excitation energy eigenvalues of $\mathcal{H}_2$ are given by E_0 and E_1, as defined in Proposition 4.7, again with multiplicities 1 and 2, respectively.*

Proof. Let $\omega = e^{i\pi/N}$. Note that E_1 is the eigenvalue corresponding to setting $k = 1$ and $x = \omega$ (or $x = 1/\omega$) in Proposition 3.3; it is easy to check that $q_1(\omega) = 0$. It suffices to show that the eigenvalues E given by (3.5) in the statement of Proposition 3.3 all satisfy $E \geq E_1$, with equality if and only if $k = 1$ and $x = \omega$ or $x = 1/\omega$, or $k = N - 1$ and $x = -\omega$ or $x = -1/\omega$.

So say $0 \leq k \leq N-1$, and $x \neq 0, \pm 1$ satisfies $q_k(x) = 0$. It suffices to show that $E \geq E_1$ for $k \leq N/2$, since $q_k(x) = 0$ implies that $q_{N-k}(-x) = 0$, and $E(k, x) = E(N - k, -x)$. Assume, therefore, that $k \leq N/2$, so that $\cos(k\pi/N) \geq 0$. Recall that either $x \in S^1$ or $x \in \mathbb{R}$.

First consider the cases $k = 0$ and $k = 1$. If $k = 0$, then $q_0(x) = (x-1)(x^{N-1}-1)$, so that x is an $(N-1)$–st root of unity, and

$$E + \frac{N}{2} - 4 = -2\left(x + \frac{1}{x}\right) \geq -4\cos\frac{2\pi}{N-1} > -2\left(1 + \cos\frac{2\pi}{N}\right) = -4\left(\cos\frac{\pi}{N}\right)^2,$$

whence $E > E_1$. If $k = 1$, then by Proposition 4.5, q_1 has no real roots besides ± 1, since $\cos\frac{\pi}{N} > \frac{N-2}{N}$ (in fact, $\cos(\pi x) > 1 - 2x$ for all $x < 1/2$). Thus $x \in S^1$. Assume that $\operatorname{Im} x > 0$; otherwise replace x by $1/x$. Then (4.3) becomes

$$\cos\frac{\pi}{N} = \frac{\sin(\left(\frac{N}{2} - 1\right)\arg x)}{\sin(\left(\frac{N}{2}\right)\arg x)}. \tag{4.4}$$

Since the right hand side of (4.4) is an increasing function of $\arg x$ for $\arg x \in (0, \pi/N)$, and since $x = e^{i\pi/N}$ satisfies (4.4), we conclude that $\arg x \in (\pi/N, \pi)$. But this gives

$$E + \frac{N}{2} - 4 = -4\left(\cos\frac{\pi}{N}\right)\cos(\arg x) \geq -4\left(\cos\frac{\pi}{N}\right)^2,$$

so that $E \geq E_1$, with equality if and only if $x = \omega$ or $x = 1/\omega$.

Now suppose that $2 \leq k \leq N/2$. If $x \in S^1$, then

$$E + \frac{N}{2} - 4 = -2\left(\cos\frac{k\pi}{N}\right)\left(x + \frac{1}{x}\right) > -2\cos\frac{k\pi}{N} \geq -2\cos\frac{2\pi}{N} = -4\left(\cos\frac{\pi}{N}\right)^2 + 2,$$

and so $E > E_1 + 2$. If $x \in \mathbb{R}$, then we may assume that $x > 0$, since otherwise $E \geq E_0 + 4$; then the following series of inequalities holds:

$$x(x-1)(x^{N-1}-1) > 0;$$

$$x + \frac{1}{x} < 1 + \frac{x^N + 1}{x^{N-1} + x} = 1 + \frac{1}{\cos\frac{k\pi}{N}} \leq \frac{1 + \cos\frac{2\pi}{N}}{\cos\frac{k\pi}{N}};$$

$$E + \frac{N}{2} - 4 = -2\left(\cos\frac{k\pi}{N}\right)\left(x + \frac{1}{x}\right) > -2\left(1 + \cos\frac{2\pi}{N}\right) = -4\left(\cos\frac{\pi}{N}\right)^2,$$

whence $E > E_1$ again. □

Corollary 4.9 *The two largest eigenvalues of each of K_1 and K_2 are $\beta_0 = 1$ and*

$$\beta_1 = 1 - \frac{4}{N}\left(\sin\frac{\pi}{N}\right)^2.$$

We conjecture that β_1 given in Corollary 4.9 is the second largest eigenvalue of K_r for all r; see Section 6.2 for partial results. We next place an upper bound on the eigenvalues of $\mathcal{H}_1$ and $\mathcal{H}_2$, or, equivalently, a lower bound on the eigenvalues of K_1 and K_2.

Proposition 4.10 *All eigenvalues E of $\mathcal{H}_1$ satisfy $E \le E_0 + 4$, while all eigenvalues E of $\mathcal{H}_2$ satisfy $E \le E_0 + 8$.*

PROOF. For $\mathcal{H}_1$, the proposition follows from Proposition 3.2. For $\mathcal{H}_2$, we claim the slightly stronger statement that $E(k,x) \le E_0 + 8$ when $x \in S^1$, and $E(k,x) \le E_0 + 2$ when $x \in \mathbb{R}$. If $x \in S^1$, this follows from Proposition 3.3. If x is real, then the inequality $E(k,x) \le E_0 + 2$ is equivalent to $\left(\cos\frac{k\pi}{N}\right)(x + 1/x) \ge 1$. By (4.2), this, in turn, is equivalent to

$$0 \le \frac{x^N + x^{N-2} + (-1)^k x^2 + (-1)^k}{x^N + (-1)^k} - 1 = \frac{x^{N-2} + (-1)^k x^2}{x^N + (-1)^k},$$

which is easy to verify for all x. □

Corollary 4.11 *All eigenvalues β of K_1 satisfy $\beta \ge 1 - 4/N$, while all eigenvalues β of K_2 satisfy $\beta \ge 1 - 8/N$. In particular, when $N \ge 8$, β_1 is the spectral gap β_* of K_1 and K_2.*

4.4 Asymptotic spectrum distribution of K_1 and K_2

In this section, we look at the asymptotic behavior of the spectra of K_1 and K_2 as $N \to \infty$, which corresponds in physics to the "classical limit." This will give additional information about the rates of convergence of these Markov chains, which will be dealt with more precisely in Section 5.3. The spectrum distributions derived below, including (4.5) and (4.6), are also derived, using the Bethe *ansatz*, by Izyumov and Skryabin [9]; their reasoning is heuristic and mathematically unsound, however, whereas we will give a rigorous derivation based on previous results in this paper.

For K_1, the limiting distribution of eigenvalues is given by Corollary 4.3 as $E = 1 - \frac{2}{N}(1 - \cos p)$, where p ranges uniformly in $[0, 2\pi)$.

From the results in Section 4.2, we may deduce the asymptotic distribution of eigenvalues of K_2 as $N \to \infty$. We first look at the limiting distribution of the bound states. There will be a bound state corresponding to k if the condition in Proposition 4.5 holds. Now if $0 \le k \le N-1$, then the inequality $\cos\left(\frac{k\pi}{N}\right) < \frac{N-2}{N}$ holds approximately when $k > 2\sqrt{N}/\pi$, and similarly $\cos\left(\frac{k\pi}{N}\right) > -\frac{N-2}{N}$ approximately when $k < N - 2\sqrt{N}/\pi$. Thus Proposition 4.5 implies that there are on the order of N bound states (there are negligibly few values of k which do not give bound states). If we write $P = 2k\pi/N$ (here $P = p_1 + p_2$ is the total momentum, where p_1 and p_2 are as in the proof of Lemma 3.4), then P is roughly uniformly distributed in $[0, 2\pi)$, and the value of $x > 1$ for which $q_k(x) = 0$ is, by (4.2), $x \approx 1/\cos(P/2)$. This leads to an eigenvalue distribution of

$$\beta = 1 - \frac{2}{N}\left(2 - \left(\cos\frac{k\pi}{N}\right)\left(x + \frac{1}{x}\right)\right) \approx 1 - \frac{1}{N}(1 - \cos P), \tag{4.5}$$

where, to reiterate, P is uniform in $[0, 2\pi)$.

We next look at the distribution of the approximately $N^2/2$ two-magnon states. These are indexed by k and x, where $q_k(x) = 0$ and $x \in S^1$; by Proposition 4.6, $\arg x$ ranges

uniformly in $[0, \pi)$ as $N \to \infty$. The relation (3.17) between $P_j = e^{ip_j}$ and k and x, from the proof of Lemma 3.4, implies that $p_1 = k\pi/N + \arg x$ and $p_2 = k\pi/N - \arg x$; in terms of p_1 and p_2, (3.12) gives us an eigenvalue distribution of

$$\beta = 1 - \frac{2}{N}((1 - \cos p_1) + (1 - \cos p_2)), \tag{4.6}$$

where we may assume that p_1 and p_2 range independently and uniformly in $[0, 2\pi)$. Note that solving for $E = N/2 - N\beta$ in this equation gives the energy distribution (2.10) from our discussion of magnons, which we previously only justified by heuristic means.

To summarize, we have found that there are approximately $N^2/2$ "two-magnon" eigenvalues of K_2, and that their distribution roughly obeys (4.6); the approximately N "bound" eigenvalues roughly obey the distribution given by (4.5).

5 Rates to stationarity

This paper culminates in the calculation, performed in this section, of rates to stationarity for the interchange walks K_1 and K_2. After describing the necessary notation and results from random walk theory in Section 5.1, we place bounds (both upper and lower) on the distance of K_1 and K_2 from stationarity after a number of steps in Sections 5.2 and 5.3, respectively. In each case, we find that it takes order N^3 steps for the chain to become random.

5.1 Background

Before we apply our Bethe *ansatz* diagonalization to the interchange random walk, we review some terminology and elementary results from the theory of random walks. A Markov chain on a finite state space X, with $|X| = n$, is characterized by a square $n \times n$ transition matrix K, so that $K(x, y)$ is the probability of ending at y after one step of the process, given initial state x. A measure $\mu : X \longrightarrow [0, 1]$ is a reversing measure for K if $\mu(x)K(x, y) = \mu(y)K(y, x)$ for all $x, y \in X$; if K has a reversing measure, then K is reversible; the unique reversing measure μ with $\sum_{x\in X} \mu(X) = 1$, which is known as the stationary distribution of K, is invariant under application of K. Finally, K is irreducible if for any $x, y \in X$, there is an m such that $K^m(x, y) > 0$. We will assume that the chain K is reversible and irreducible, so that the matrix K is real symmetric; then K is completely diagonalizable, with eigenvalues $1 = \beta_0 > \beta_1 \geq \beta_2 \geq \ldots \geq \beta_{N-1} > -1$.

The reversing measure μ defines an inner product space $L^2(\mu)$ of complex-valued functions on X, where the inner product is given by $\langle f|g\rangle = \sum_{x\in X} f(x)\overline{g(x)}\mu(x)$, and the L^2 norm on $L^2(\mu)$ is defined by $\|f\|_2 = \langle f|f\rangle^{1/2}$. Probability distributions on X can then be considered as elements in $L^2(\mu)$. Note that if $\mu(x) \equiv 1/n$ is uniform, then the $L^2(\mu)$ inner product $\langle\cdot|\cdot\rangle$ is related to the usual hermitian inner product $(\cdot, \cdot)$ by $n\langle f|g\rangle = (f, g)$.

We wish to measure the distance of a probability distribution on X from stationarity. In particular, let $K^m_x(y) = K^m(x, y)$ be the (x, y) entry in the matrix K^m, or the probability that the chain will be in state y after m steps, given that it begins in state x. The more "random" the distribution K^m_x is, the closer it will be to stationarity. To quantify the distance between a distribution and stationarity, we use the total variation distance

$$\|K^m_x - \mu\|_{\mathrm{TV}} = \max_{A\subset X} |K^m_x(A) - \mu(A)| = \frac{1}{2} \max_{\|f\|_\infty \leq 1} |K^m_x(f) - \mu(f)|, \tag{5.1}$$

where $P(A)$ denotes $\sum_{x\in A} P(x)$, $P(f)$ denotes $\sum_{x\in X} f(x)P(x)$, and $\|f\|_\infty = \max_{x\in X} |f(x)|$. The following proposition, essentially taken from Diaconis and Saloff-Coste [4], summarizes some elementary results from random walk theory, used to obtain upper bounds on distances from stationarity.

Proposition 5.1 *Let $\{\varphi_j\}_{j=0}^{n-1}$ be an orthonormal basis of eigenvectors of K in $L^2(\mu)$, ordered to satisfy $K\varphi_j = \beta_j\varphi_j$, where $1 = \beta_0 > \beta_1 \geq \cdots \geq \beta_{n-1} > -1$ are the eigenvalues of K; also, let $\|\cdot\|_2$ denote the $L^2(\mu)$ norm, and let $\beta_* = \max\{\beta_1, |\beta_{n-1}|\}$ be the spectral gap*

of K. Then we have

$$\|K_x^m - \mu\|_{\mathrm{TV}} \le \frac{1}{2} \left\| \frac{K_x^m}{\mu} - 1 \right\|_2 = \frac{1}{2} \left(\sum_{j=1}^{n-1} |\varphi_j(x)|^2 \beta_j^{2m} \right)^{1/2}, \tag{5.2}$$

$$\|K_x^m - \mu\|_{\mathrm{TV}} < \frac{\sqrt{n}}{2} \beta_*^m, \tag{5.3}$$

$$\frac{1}{n} \sum_{x \in X} \|K_x^m - \mu\|_{\mathrm{TV}}^2 \le \frac{1}{4} \sum_{j=1}^{n-1} \beta_i^{2m}. \tag{5.4}$$

The proof of Proposition 5.1 is standard; see, e.g., Diaconis and Saloff-Coste [4]. The inequality given by (5.3) is usually somewhat crude, since it only uses one eigenvalue rather than all of them. Inequality (5.4) states that the average (root-mean-square) distance from stationarity, averaged over all initial states, is bounded above by $(\sum_{j=1}^{n-1} \beta_j^{2m})^{1/2}/2$; when K is vertex transitive, as is the case for the interchange process K_1, all distances from stationarity are the same, and we obtain $\|K_x^m - \mu\|_{\mathrm{TV}} \le (\sum_{j=1}^{n-1} \beta_j^{2m})^{1/2}/2$ for all initial states x, which is usually better than (5.3) by a factor of $\sqrt{n}$.

To find a lower bound on distance from stationarity, we have the following useful result.

Proposition 5.2 *If $\mu \equiv 1/n$ is uniform and $K\varphi = \beta\varphi$, then*

$$\|K_x^m - \mu\|_{\mathrm{TV}} \ge \frac{|\varphi(x)|}{2\|\varphi\|_\infty} \beta^m,$$

where $\|\varphi\|_\infty = \max |\varphi(x)|$.

Proof. If $\beta = 1$, the result is trivial. Otherwise, φ is orthogonal to the constant eigenvector of K, so that $\mu(\varphi) = 0$. Introduce δ_x defined by $\delta_x(x) = 1/\mu(x)$, $\delta_x(y) = 0$ if $y \ne x$. Then

$$K_x^m(\varphi) = \langle \varphi | K^m \delta_x \rangle = \langle K^m \varphi | \delta_x \rangle = \beta^m \langle \varphi | \delta_x \rangle = \beta^m \varphi(x).$$

We now use the definition (5.1) of total variation distance, with $f = \varphi/\|\varphi\|_\infty$:

$$\|K_x^m - \mu\|_{\mathrm{TV}} \ge \frac{1}{2} |K_x^m(f) - \mu(f)| = \frac{|\varphi(x)|}{2\|\varphi\|_\infty} \beta^m. \quad \square$$

5.2 Rate to stationarity of K_1

As mentioned in Section 4.1, K_1 is essentially the simple random walk on $\mathbb{Z}_N$, but with positive holding probability. To bound rates of convergence of K_1, we wish to bound the sum of powers of the eigenvalues of K_1, by the results of the previous section.

Lemma 5.3 *Let $\{\beta_k : 0 \le k \le N-1\}$ be the set of eigenvalues of K_1, as given in Corollary 4.3, and let $m > N^3$. Then*

$$\sum_{k=1}^{N-1} \beta_k^m < 3e^{-4\pi^2 m/N^3 + \gamma m/N^5},$$

where $\gamma = 4\pi^4/3$.

PROOF. Taylor series expansion reveals that $1 - \cos x \ge x^2/2 - x^4/24$ for any $x \in [0, \pi]$, so that

$$1 - \alpha(1 - \cos x) \le e^{-\alpha(x^2/2 - x^4/24)} \tag{5.5}$$

for any $\alpha > 0$ and $x \in [0, \pi]$. It follows that if $k \le N/2$, then $\beta_k \le e^{-4k^2\pi^2/N^3 + \gamma k^4/N^5} \le e^{-4k\pi^2/N^3 + \gamma/N^5}$; the first inequality follows from (5.5), and the second is straightforward to check. Thus we have

$$\sum_{k=1}^{N-1} \beta_k^m \le 2 \sum_{k=1}^{\lfloor N/2 \rfloor} \beta_k^m < 2 \sum_{k=1}^{\infty} e^{-4k\pi^2 m/N^3 + \gamma m/N^5} < 3e^{-4\pi^2 m/N^3 + \gamma m/N^5},$$

where the final inequality assumes that $m > N^3$. □

We are now in a position to place extremely sharp upper and lower bounds on the distance from stationarity of $(K_1)_x^m$, the probability distribution after m steps, given initial state x. Proposition 5.4 shows that it takes order N^3 steps for the chain K_1 to become random. This is more or less expected, since K_1 runs $N/2$ times as slowly as the simple random walk on $\mathbb{Z}_N$, which takes order N^2 steps to become random; more specifically, if Q represents the simple random walk, then $\|Q^m - \mu\| \sim e^{-\pi^2 m/2N^2}$ [3, §3C].

Proposition 5.4 *Let x be any initial state of K_1. If $m > N^3$, then*

$$\|(K_1)_x^m - \mu\|_{\mathrm{TV}} \le \tfrac{3}{2} e^{-4\pi^2 m/N^3 + \gamma m/N^5}, \tag{5.6}$$

where $\gamma = 4\pi^4/3$; conversely, for any m,

$$\|(K_1)_x^m - \mu\|_{\mathrm{TV}} \ge \tfrac{1}{2} e^{-4\pi^2 m/N^3}. \tag{5.7}$$

PROOF. By equation (5.4), Lemma 5.3, and the fact that K_1 is vertex transitive on its state space $\mathbb{Z}_N$, we deduce the upper bound (5.6). The lower bound (5.7) is a straightforward application of Proposition 5.2, with $\beta = \beta_1 = 1 - \frac{4}{N}\left(\sin \frac{\pi}{N}\right)^2$ and corresponding eigenvector φ defined by $\varphi(j) = e^{2i\pi j/N}$; we have $\|\varphi\|_\infty = 1$ and $|\varphi(x)| = 1$ for all x. Now the inequality $1 - \alpha(\sin x)^2 \ge e^{-\alpha x^2}$, which holds for all $\alpha \ge 0$ and $x \in \mathbb{R}$ and is easily shown by calculus, implies, when $\alpha = 4/N$ and $x = \pi/N$, that $\beta_1 \ge e^{-4\pi^2/N^3}$; the proposition follows. □

5.3 Rate to stationarity of K_2

In accordance with the results from Section 5.1, we wish to bound $\sum \beta(k,x)^m$, where m is an integer and the sum is over all eigenvalues $\beta(k,x) \neq 1$ of K_2. Since the second largest eigenvalue of K_2, by Corollary 4.9, is approximately $1-4\pi^2/N^3 \approx e^{-4\pi^2/N^3+4\pi^4/3N^5}$, this sum is roughly bounded below by $e^{-4\pi^2 m/N^3+4\pi^4/3N^5}$. We will show that the sum is also bounded above, up to constants, by this exponential. The condition $m > N^3$ in the proposition can be replaced by a much weaker condition on m, although (5.8) does not hold for, say, $m = 1$ and N large; in any case, our interest lies in the behavior of $\sum \beta(k,x)^m$ when m is large, and so $m > N^3$ is an acceptable condition.

Lemma 5.5 *If $m > N^3$, then*

$$\sum_{\beta(k,x)\neq 1} \beta(k,x)^m < 8e^{-4\pi^2 m/N^3+\gamma m/N^5}, \tag{5.8}$$

where $\gamma = 4\pi^4/3$.

This result is, up to constants, more or less intuitively clear; the proof is rather involved and not at all illuminating, but is included for completeness. We will need the following simple lemma.

Lemma 5.6 *If μ and ν are real numbers with $2\pi/N \leq \mu+\nu \leq \pi$, then*

$$1-\cos\mu\cos\nu > \frac{(\mu+\nu)\pi}{2N} - \frac{\pi^4}{3N^4}.$$

Proof. It is straightforward to check that the function $y^2 - \pi y/N - y^4/3$ takes on its minimum value over the interval $[\pi/N, \pi/2]$ when $y = \pi/N$, so that $y^2 - y^4/3 \geq \pi y/N - \pi^4/3N^4$ on this interval. But Taylor expansion yields $(\sin x)^2 > y^2 - y^4/3$, and so

$$1-\cos\mu\cos\nu \geq 1-\left(\cos\frac{\mu+\nu}{2}\right)^2 = \left(\sin\frac{\mu+\nu}{2}\right)^2 > \frac{(\mu+\nu)\pi}{2N} - \frac{\pi^4}{3N^4}. \quad \square$$

Proof of Lemma 5.5. For brevity, we omit some steps in the proof. It will be convenient to split the sum into two sums, one over eigenvalues from the "two-magnon" states, and the other over eigenvalues from the "bound" states. We first bound the sum over "two-magnon" eigenvalues, i.e., those eigenvalues $\beta(k,x)$ from Proposition 3.3 with $x \in S^1$. Since the roots of q_{N-k} give rise to the same energy eigenvalues as the roots of q_k, the sum of $\beta(k,x)^m$ over all k and $x \in S^1$ (roots of $q_k(x)$) is at most twice the same sum over $k \leq N/2$.

Let $k \leq N/2$. Consider the roots of q_k in S^1 with positive imaginary part; these may be labelled $x_{k,0}, \ldots, x_{k,\ell}$ ($\ell \approx N/2$ depends on k) in increasing order of argument, so that $0 < \arg x_{k,0} < \cdots < \arg x_{k,\ell} < \pi$. Write $\theta_{k,j} = \arg x_{k,j}$, and write $\xi_{k,j} = k + N\theta_{k,j}/\pi$. We now bound $\xi_{k,j}$. When $k = 0$, we may directly calculate that $\xi_{0,j} = N\theta_{0,j}/\pi = \frac{2(j+1)N}{N-1} \geq j+2$. For other values of k, Proposition 4.6 implies bounds on $\xi_{k,j}$ as follows: $\xi_{1,j} \geq j+2$; $\xi_{2,j} \geq j+3$; and $\xi_{k,j} \geq j+k$ for $k \geq 3$.

In all cases, $2 \leq \xi_{k,j} \leq N$, and so we may apply Lemma 5.6 to conclude that

$$\beta(k, x_{k,j}) = 1 - \frac{4}{N}(1 - \cos(k\pi/N)\cos\theta_{k,j}) > 1 - \frac{2\xi_{k,j}\pi^2}{N^3} + \frac{4\pi^4}{3N^5} \geq e^{4\pi^4/3N^5}(e^{2\pi^2/N^3})^{\xi_{k,j}}.$$

Raising this inequality to the m-th power, summing over all two-magnon eigenvalues $\beta(k, x_{k,j})$ with $k \leq N/2$, and estimating the resulting sum by an infinite sum via the bounds on $\xi_{k,j}$ gives

$$\sum_{k \leq N/2,\, x \in S^1} \beta(k,x)^m < \left(\frac{2\eta^{2m} - \eta^{4m}}{(1-\eta^m)^2}\right) e^{\gamma m/N^5} < 3\eta^{2m} e^{\gamma m/N^5},$$

where η denotes $e^{-2\pi^2/N^3}$, and we have assumed that $m > N^3$ (so that $\eta^m < e^{-2\pi^2}$).

We next wish to sum $\beta(k,x)^m$ over all k and $x \in \mathbb{R} - \{0, \pm 1\}$ satisfying $q_k(x) = 0$. As before, we bound this above by twice the same sum over $k \leq N/2$. Now suppose that $k \leq N/2$, and that $x > 1$ satisfies $q_k(x) = 0$; we consider two cases. If k is even, then by (4.2) and the fact that $\frac{x^{N-1}+x}{x^N+1} > 1/x$, we have that $x > 1/\cos(k\pi/N)$, so that $x + 1/x < \cos(k\pi/N) + 1/\cos(k\pi/N)$, and Corollary 4.4 gives

$$\beta(k,x) < 1 - \frac{2}{N}\left(\sin\frac{k\pi}{N}\right)^2 \leq e^{-2k\pi^2/N^3 + \gamma/2N^5} \leq \eta^{2k} e^{\gamma/N^5},$$

as in the proof of Lemma 5.3.

If k is odd, then denote $\frac{N-2}{N}$ by c; by Proposition 4.5, $\cos(k\pi/N) < c$. Now by (4.2) and the fact that $\frac{x^{N-1}-x}{x^N-1} > c/x$, we have that $x + 1/x < \cos(k\pi/N)/c + c/\cos(k\pi/N)$, and so

$$\beta(k,x) < 1 - \frac{2}{N}\left(2 - c - \frac{(\cos\frac{k\pi}{N})^2}{c}\right) < 1 - \frac{4(1-c)}{N} = 1 - \frac{8}{N^2} < e^{-8/N^2}.$$

We find that

$$\sum_{k \leq N/2,\, x > 1} \beta(k,x)^m < e^{\gamma m/N^5}\left(\frac{\eta^{4m}}{1-\eta^{2m}} + Ne^{-8m/N^2}\right) < \eta^{2m} e^{\gamma m/N^5}.$$

The last inequality is a crude estimate which definitely holds when $m > N^3$. So the sum of $\beta(k,x)^m$ over all bound states is as most twice this, or $2\eta^{2m}$. Combining this with the $6\eta^{2m}$ upper bound for the sum of $\beta(k,x)^m$ over all two-magnon states gives an upper bound of $8\eta^{2m}e^{\gamma m/N^5} = 8e^{-4\pi^2 m/N^3 + \gamma m/N^5}$ for the total sum $\sum \beta(k,x)^m$.

When N is even, we must add the eigenvalue $\beta = 1 - 2/N$; but we may absorb $(1-2/N)^m$ into the bound state sum, and the same upper bound holds. □

Proposition 5.1 and Lemma 5.5 immediately give the following result.

Proposition 5.7 *If $m > N^3$ and $\gamma = 4\pi^4/3$, then*

$$\frac{1}{\binom{N}{2}} \sum_{1 \leq x_1 < x_2 \leq N} \|(K_2)^m_{|x_1 x_2\rangle} - \mu\|_{\mathrm{TV}} < 2e^{-4\pi^2 m/N^3 + \gamma m/N^5}.$$

In other words, it takes order N^3 steps, on average, for K_2 to become random. Proposition 5.1 and the fact that $\beta_* \leq e^{-4\pi^2/N^3+\gamma/N^5}$ from the proof of Lemma 5.5 also imply a weak upper bound on individual total variation distances:

$$\|(K_2)^m_{|x_1x_2\rangle} - \mu\|_{\text{TV}} < 4Ne^{-4\pi^2 m/N^3+\gamma m/N^5}.$$

We conjecture that the "correct" upper bound is on the order of $e^{-4\pi^2 m/N^3+\gamma m/N^5}$ for all initial states $|x_1x_2\rangle$.

To complement the upper bound on distance from stationarity provided by Proposition 5.7, we next give the corresponding lower bound.

Proposition 5.8 *For any m,*

$$\|(K_2)^m_{|x_1x_2\rangle} - \mu\|_{\text{TV}} \geq \frac{1}{2}\left|\cos\left(\frac{\pi(x_2-x_1)}{N}\right)\right| e^{-4\pi^2 m/N^3}.$$

PROOF. We use Proposition 5.2, with $\beta = \beta(1, e^{i\pi/N})$, the second largest eigenvalue of K_1. The proof of Lemma 3.4 constructs an eigenvector ψ of K_1 corresponding to eigenvalue $\beta(1, e^{i\pi/N})$, given by

$$\psi(x_1, x_2) = a(x_1, x_2) = e^{2i\pi x_1/N} + e^{2i\pi x_2/N}.$$

Clearly $\|\psi\|_\infty \leq 2$, and $|\psi(x_1, x_2)| = 2\left|\cos\left(\frac{\pi(x_2-x_1)}{N}\right)\right|$; from the proof of Proposition 5.4, we also know that $\beta = 1 - \frac{4}{N}\left(\sin\frac{\pi}{N}\right)^2 \leq e^{-4\pi^2/N^3}$, and the result follows. □

If N is even, then Proposition 5.8 is vacuous when $x_2 - x_1 = N/2$; in this case, using Proposition 5.2 with $\beta = \beta(2, e^{2i\pi/N})$ gives the lower bound $\|K_x^m - \mu\|_{\text{TV}} \geq e^{-16\pi^2 m/N^3}/2$.

6 Other walks

This section is devoted to Markov chains other than K_1 and K_2. Section 6.1 introduces the exclusion model, and demonstrates how our results from previous sections may be used to improve on currently known bounds on the eigenvalues of the exclusion model on $\mathbb{Z}_N$. Section 6.2 gives partial results towards an analysis of the general interchange model K_r. In Section 6.3, we turn the tables and show how random walk theory may be applied to physics and, in particular, statistical mechanics.

6.1 Exclusion model

We may apply our work on the interchange model to other chains as well. One simple application involves the exclusion model; we describe this model, deduce bounds on its convergence, and compare our bounds to the best previously known bounds, obtained through comparison with the Bernoulli-Laplace model of diffusion.

Given a regular undirected graph G, the (symmetric) exclusion model on G has as its state space all $\binom{N}{r}$ possible placings of r indistinguishable objects on the vertices of G, where N is the number of vertices in G, and $r < N$ is fixed. A step in the exclusion process constitutes choosing one of the objects uniformly at random, choosing one of its neighbors in the graph uniformly at random, and either moving the object to the neighboring vertex if it is unoccupied, or leaving the system as is, if it is occupied. Formally, let $V(G)$ and $E(G)$ be the vertex and edge sets of G, respectively, and let d be the common degree of all vertices in G. The state space is all r-subsets of $V(G)$, and the transition probability from A_1 to A_2 is

$$P_r(A_1, A_2) = \begin{cases} 0 & \text{if } |A_1 \cap A_2| \leq r-2 \\ 0 & \text{if } |A_1 \cap A_2| = r-1 \text{ and } A_1 - A_2 = \{v_1\}, \\ & \quad A_2 - A_1 = \{v_2\}, \text{ with } (v_1, v_2) \notin E(G) \\ 1/(rd) & \text{if } |A_1 \cap A_2| = r-1 \text{ and } A_1 - A_2 = \{v_1\}, \\ & \quad A_2 - A_1 = \{v_2\}, \text{ with } (v_1, v_2) \in E(G) \\ \frac{\sum_{v \in A_1} d(v, A_1)}{rd} & \text{if } A_1 = A_2, \text{ where} \\ & \quad d(v, A_1) = |\{v' \in A_1 : (v, v') \in E(G)\}|, \end{cases} \tag{6.1}$$

when A_1 and A_2 are r-subsets of $V(G)$. For extensive background on the exclusion model and its significance, see Liggett [12].

We wish to consider the exclusion model on a cycle on N vertices, also known as the (undirected) Cayley graph of $\mathbb{Z}_N$ with generators $\{\pm 1\}$; let P_r be the corresponding transition matrix. Like the Heisenberg XXX Hamiltonian $\mathcal{H}$, the exclusion matrix P_r is linearly related to the interchange matrix K_r; this is the content of the following observation.

Proposition 6.1 $K_r = 1 - \frac{2r}{N}(1 - P_r)$.

Proof. Compare the definitions (4.1) and (6.1) of K_r and P_r. If $2r \leq N$, then the interchange process may be described as performing the exchange process with probability $2r/N$, and doing nothing with probability $1 - 2r/N$; the result follows. If $2r > N$, then

exchange is the same as interchange with probability $N/2r$, and holding with probability $1 - N/2r$. $\square$

Along the same lines, we may prove more generally that a generalized interchange model K'_r on a regular graph with common vertex degree d, in which there are r red cards and $N-r$ cards, and each move chooses two neighboring vertices at random and interchanges them, is related to the corresponding exchange model P'_r by $K'_r = 1 - \frac{rd}{N}(1 - P'_r)$.

Propositions 4.9 and 6.1 imply that the spectral gap of each of P_1 and P_2 is $\lambda_1 = 1 - \left(\sin \frac{\pi}{N}\right)^2$; we believe that the precise calculation of the spectral gap for P_2 is new. As for the interchange model, the spectral gap eigenvalue dominates the sum (5.4) of powers of eigenvalues, and we conclude that it takes, on average, about N^2/π^2 steps for each of P_1 and P_2 to reach stationarity:

$$\sum_{\lambda \neq 1} \lambda^m < 8e^{-\pi^2 m/N^2 + \pi^4 m/3N^4}, \tag{6.2}$$

where the sum is over eigenvalues λ of P_1 or P_2. (A precise proof of this statement requires a series of involved estimates nearly identical to those presented in Section 5.3, and is omitted here.)

The best previously known approach for bounding time to stationarity in the exclusion model is the method of comparison, as described in Diaconis and Saloff-Coste [4]. This technique bounds the eigenvalues of the exclusion matrix P_r by comparing the exclusion model with another Markov chain, the classical Bernoulli-Laplace model of diffusion. Through Proposition 6.1, comparison can be used to bound the eigenvalues of the interchange matrix K_r for any r, thus providing more general results than our Bethe *ansatz* approach, which is limited to $r = 1$ and $r = 2$. In the case $r = 2$, however, the Bethe *ansatz* method gives the correct bound on convergence of the interchange and exclusion models on $\mathbb{Z}_N$, while the comparison method, as we shall see, is off by a $\log N$ term.

Before discussing the results of comparison, we briefly describe the Bernoulli-Laplace model of diffusion; see Feller [7] for applications of this model. In Bernoulli-Laplace diffusion, there are r objects on one side of a partition, and $N - r$ on the other; each step consists of choosing one object from each side, and switching them. Note that the Bernoulli-Laplace chain is linearly related to the generalized interchange model on $K_{r,N-r}$, the bipartite graph on groups of r and $N - r$ vertices; interchange on $K_{r,N-r}$ is equivalent to Bernoulli-Laplace with probability $r(N-r)/\binom{N}{2}$ and holding with probability $1-r(N-r)/\binom{N}{2}$. The eigenvalues of the Bernoulli-Laplace matrix $\tilde{P}_r$ are known and relatively simple (unlike those of the interchange model on $\mathbb{Z}_N$); Diaconis and Shahshahani [5] compute them to be $1 - \frac{j(N-j+1)}{r(N-r)}$, $0 \leq j \leq r$, with multiplicity $\binom{N}{j} - \binom{N}{j-1}$.

A detailed discussion of comparison techniques is beyond the scope of this paper; the interested reader is referred to Diaconis and Saloff-Coste [4]. For our purposes, it suffices to note that we can bound the eigenvalues of the exclusion process by comparison with the Bernoulli-Laplace model via the following result of Diaconis and Saloff-Coste, which is a special case of Theorem 3.1 in [4].

Proposition 6.2 *The eigenvalues* $1 = \lambda_0 > \lambda_1 \geq \cdots \geq \lambda_{\binom{N}{r}-1}$ *of the exclusion process* P_r *are bounded above by*

$$\lambda_i \leq 1 - \frac{k_i(N - k_i + 1)}{2r\Delta_0},$$

where $k_i = j$ *when* $\binom{N}{j-1} \leq i < \binom{N}{j}$, $0 \leq j \leq \min\{r, N-r\}$, *and* $\Delta_0 = m(m+1)(2m+1)/6$, $m = \lfloor N/2 \rfloor$.

In the case $r = 2$, Proposition 6.2 implies that, besides the trivial eigenvalue 1, $N-1$ of the eigenvalues of P_2 are bounded above by $1 - \frac{6N}{(N+1)^3}$, and the other $\binom{N}{2} - N$ eigenvalues are bounded above by $1 - \frac{12(N-1)}{(N+1)^3}$. (In fact, the second largest eigenvalue λ_1 of P_2 is approximately $1 - \pi^2/N^2$, by Propositions 4.9 and 6.1, so that the upper bound of $1 - \frac{6N}{(N+1)^3}$ for λ_1 provided by comparison is quite close.) Comparison thus implies that the sum of the m-th powers of the eigenvalues of K_2 is bounded above as follows:

$$\sum_{\lambda \neq 1} \lambda^m < \binom{N}{2} \left(1 - \frac{6N}{(N+1)^3}\right)^m \sim \tfrac{1}{2} e^{-6m/N^2 + 2\log N}. \tag{6.3}$$

This upper bound is off by a $\log N$ term, compared to our result (6.2), obtained via the Bethe *ansatz*. Since the upper bound on every eigenvalue is on the order of $1 - c/N^3$, the upper bound on the sum of their powers cannot be much improved from (6.3); the best upper bound from comparison methods will contain a $\log N$ term.

6.2 General interchange model $\mathcal{H}_r$

We next consider the general interchange model $\mathcal{H}_r$ on $\mathbb{Z}_N$. As noted in Section 3.2, solving for the exact eigenvalues of $\mathcal{H}_r$ is equivalent to solving a many body problem, and is thus essentially impossible. We can, however, find some of the eigenvalues of $\mathcal{H}_r$; our main result in this section, Proposition 6.5, states that if $r < N/2$, then $\mathcal{H}_{r+1}$ inherits all of the eigenvalues of $\mathcal{H}_r$. Proving this is more difficult than it may appear, and requires the following purely combinatorial result.

Lemma 6.3 *Let* r *be an integer, and, for any* k *between 0 and* $2r+1$, *let* V_k *be the* $\binom{2r+1}{k}$*-dimensional vector space, over a field of characteristic 0, formally generated by the basis* $\{v_A : A \subset \{1, \ldots, 2r+1\},\ |A| = k\}$. *Then the map* $\varphi : V_r \longrightarrow V_{r+1}$ *defined by*

$$\varphi(v_A) = \sum_{B \supset A,\, |B| = r+1} v_B = \sum_{1 \leq k \leq 2r+1,\, k \notin A} v_{A \cup \{k\}}$$

is an isomorphism.

Proof. We explicitly construct an inverse to φ as follows:

$$\varphi^{-1}(v_B) = \sum_{|A| = r} f(|A - B|)\, v_A, \tag{6.4}$$

where f is some function on $\{0, 1, \ldots, r\}$ to be determined. If $|C| = r + 1$, then we want the coefficient of v_C in $\sum_{|A|=r} f(|A - B|)\, \varphi(v_A)$ to be 1 if $C = B$, and 0 otherwise; this will imply that (6.4) gives a formula for φ^{-1}. To calculate this coefficient, let $|A - C| = |C - A| = m$. Note that there are r sets A with $|A| = r$ and $A \subset C$, and these are exactly the sets for which $\varphi(v_A)$ will contain a nonzero coefficient of v_C; of these, m satisfy $|A - B| = m - 1$, while the other $r + 1 - m$ satisfy $|A - B| = m$. Thus the coefficient in question is $mf(m-1) + (r+1-m)f(m)$ (when $1 \le m \le r+1$; the coefficient is $(r+1)f(0)$ when $m = 0$), and (6.4) will give a formula for φ^{-1} if and only if $f(0) = 1/(r + 1)$ and $mf(m - 1) + (r + 1 - m)f(m) = 0$ for $1 \le m \le r + 1$. But this is satisfied by

$$f(m) = \frac{(-1)^m}{(r+1)\binom{r}{m}}, \tag{6.5}$$

and we conclude that φ is indeed an isomorphism, with inverse given by (6.4) and (6.5). □

To use the above lemma, we translate some of the physics notation from Section 2.2 into combinatorial language. Identify the vector $|x_1, \ldots, x_r\rangle$ with the formal symbol $v_{\{x_1,\ldots,x_r\}}$, so that v_A is the vector corresponding to the state in which down spins (or red cards) are precisely in the positions corresponding to the elements of A. Then $\{v_A : A \subset \mathbb{Z}_N,\ |A| = r\}$ is the standard basis of $\mathfrak{H}_r$. Now consider the operator σ^-_{tot}, defined in Section 2.4 as $\sum_{j=1}^N \sigma^-_j$. Since σ^-_j sends an up spin in the j-th position to a down spin, and a down spin in the j-th position to zero, it is easy to see that σ^-_{tot} may be defined in our combinatorial language as follows:

$$\sigma^-_{\text{tot}}(v_A) = \sum_{k \in \mathbb{Z}_N,\, k \notin A} v_{A \cup \{k\}},$$

where A is any subset of $\mathbb{Z}_N$. In this form, it is clear that σ^-_{tot} maps $\mathfrak{H}_r$ to $\mathfrak{H}_{r+1}$ when $0 \le r < N$; in physics language, σ^-_{tot} raises the number of down spins by one.

Lemma 6.4 *If $r < N/2$, then the collection of vectors $\{\sigma^-_{\text{tot}} v_A : A \subset \mathbb{Z}_N,\ |A| = r\}$ is linearly independent.*

Proof. Suppose that constants α_A satisfy $\sum_{|A|=r} \alpha_A \sigma^-_{\text{tot}} v_A = 0$. Let C be any subset of $\mathbb{Z}_N$ with $|C| = 2r + 1$. For any $B \subset C$ with $|B| = r + 1$, the coefficient of v_B in the above sum is the sum of α_A over all A with $A \subset B$ and $|A| = r$; but this is precisely the coefficient of v_B in

$$\varphi\left(\sum_{A \subset C,\, |A|=r} \alpha_A v_A\right), \tag{6.6}$$

where φ is as in the statement of Lemma 6.3, with C replacing $\{1, \ldots, 2r + 1\}$. Since this coefficient must be 0 for all $B \subset C$ with $|B| = r + 1$, we conclude that (6.6) must be equal to 0. But φ is injective by Lemma 6.3, and therefore $\alpha_A = 0$ for all $A \subset C$. This is true for any $C \subset \mathbb{Z}_N$ with $|C| = 2r + 1$, so that $\alpha_A = 0$ for all A, and the result follows. □

Proposition 6.5 *If $r < N/2$, then the multiset of eigenvectors of $\mathcal{H}_r$ (respectively K_r) is contained in the multiset of eigenvectors of $\mathcal{H}_{r+1}$ (respectively K_{r+1}).*

Proof. By Corollary 4.2, the result for K follows from the result for $\mathcal{H}$, and so it suffices to prove the proposition for $\mathcal{H}$. Now σ_{tot}^- commutes with $\mathcal{H}$; this may be proven directly from the definition of $\mathcal{H}$ and the easily computed commutation relations $[\sigma_j^-, \sigma_j^x] = i\sigma_j^z$, $[\sigma_j^-, \sigma_j^y] = -i\sigma_j^z$, $[\sigma_j^-, \sigma_j^z] = -i\sigma_j^-$. Thus any eigenvector ψ of $\mathcal{H}_r$ gives rise to an eigenvector $\sigma_{\text{tot}}^-\psi$ of $\mathcal{H}_{r+1}$ with the same eigenvalue. Now Lemma 6.4 states that σ_{tot}^- maps the standard basis of $\mathcal{H}_r$ to a linearly independent set in $\mathcal{H}_{r+1}$; it is a simple exercise in linear algebra to conclude that σ_{tot}^- maps any basis of $\mathcal{H}_r$ to a linearly independent set in $\mathcal{H}_{r+1}$. In particular, a basis of eigenvectors of $\mathcal{H}_r$ is mapped to a linearly independent collection of eigenvectors of $\mathcal{H}_{r+1}$ with the same eigenvalues, and the proposition follows. □

As an illustration of Proposition 6.5, note that the eigenvalue β_k of K_1 corresponds to the eigenvalue $\beta(k, e^{i\pi k/N})$ of K_2, a "two-magnon" eigenvalue.

Corollary 6.6 *If $r < N/2$, then the spectral gap of K_{r+1} is no smaller than the spectral gap of K_r.*

We conjecture that the spectral gap of K_r is in fact the same for all r; this is true for $r = 1$ and $r = 2$, by Corollary 4.11. The treatment of the Heisenberg antiferromagnet, given by equation (2.4) with $J_x = J_y = J_z = -1$, by Hulthén [8] shows that the ground state energy of the antiferromagnet $\mathcal{H}$ is asymptotically $N/2 - 2N\log 2$; this means that the highest energy eigenvalue of the corresponding XXX ferromagnet is $-N/2 + 2N\log 2$, so that the smallest eigenvalue of K_r is at least $1 - 2\log 2 \approx -0.39$, and therefore the spectral gap of K_r is in fact the second largest eigenvalue β_1. The best general result in the direction of this conjecture is given by the comparison techniques mentioned in Section 6.1. For general r, Proposition 6.2 implies the following upper bound on the second largest eigenvalue β_1 of the interchange chain K_r, complemented by a lower bound implied by Proposition 6.5:

$$1 - \frac{4\pi^2}{N^3} \leq \beta_1 < 1 - \frac{24}{(N+2)^3}.$$

The conjecture is that equality holds for the lower bound.

Proposition 6.5 has an interpretation in terms of representation theory; we briefly describe this, assuming familiarity with some basic representation theory, on the level of Diaconis [3]. The state space X_r of K_r consists of the collection of r-subsets of $\{1, \ldots, N\}$, corresponding to the locations of the red cards. Since the elements of X_r generate the vector space $\mathfrak{H}_r$, there is an obvious isomorphism between $\mathfrak{H}_r$ and $L(X_r)$, the vector space of complex valued functions on X_r. Recall from Section 4.1 that X_r is itself identical to $S_N/(S_r \times S_{N-r})$. Then the group S_N acts on $L(X_r)$ through its action on X_r, and it follows that $L(X_r) \cong \mathfrak{H}_r$ gives a representation on S_N. Following Diaconis [3], we will denote this representation by $M^{(N-r,r)}$; James [10] writes it as $[N-r][r]$. (The special case $M^{(N-1,1)}$ is sometimes called the natural representation of S_N.)

Assume that $2r \leq N$. Then $M^{(N-r,r)}$ has a decomposition into irreducible representations of S_N, which, by Young's rule (see [10]), is given by

$$\mathfrak{H}_r \cong M^{(N-r,r)} \cong S^N \oplus S^{(N-1,1)} \oplus S^{(N-2,2)} \oplus \cdots \oplus S^{(N-r,r)}, \tag{6.7}$$

where $S^{(N-k,k)}$ is the so-called Specht module associated to the partition $(N-k,k)$ of N. Thus $M^{(N-r-1,r+1)} \cong M^{(N-r,r)} \oplus S^{(N-r-1,r+1)}$, and so $\mathfrak{H}_r$ is a direct summand of $\mathfrak{H}_{r+1}$ when $2r < N$. But this is precisely what we proved in Proposition 6.5 (or, more precisely, in Lemma 6.4): the map σ^-_{tot} is an inclusion from $\mathfrak{H}_r$ to $\mathfrak{H}_{r+1}$, and so $\mathfrak{H}_{r+1} \cong \mathfrak{H}_r \oplus (\mathfrak{H}_r)^\perp$. Writing $(\mathfrak{H}_r)^\perp$ as $S^{(N-r-1,r+1)}$ (with $S^N = \mathfrak{H}_0$), we recover (6.7).

In terms of eigenvectors, the $\binom{N}{r}$-dimensional space $\mathfrak{H}_r$ inherits $\binom{N}{r-1}$ eigenvectors of $\mathcal{H}$ from $\mathfrak{H}_{r-1}$; this collection of eigenvectors of $\mathcal{H}$ in $\mathfrak{H}_r$ is completed by a basis of $(\mathfrak{H}_{r-1})^\perp \subset \mathfrak{H}_r$ consisting of another $\binom{N}{r} - \binom{N}{r-1}$ eigenvectors of $\mathcal{H}$. The Specht module $S^{(n-r,r)}$ is generated by eigenvectors of $\mathcal{H}$ in $\mathfrak{H}_r$ not derived (via σ^-_{tot}) from $\mathfrak{H}_{r-1}$.

More concretely, view $\mathfrak{H}_r$ as $L(X_r)$, the space of complex valued functions on r-subsets of $\{1,\ldots,N\}$. Then S^N in the decomposition (6.7) of $\mathfrak{H}_r$ is the one-dimensional subspace of constant functions on X_r; this corresponds, in the notation of physics, to $\sum_{1\leq x_1<\cdots<x_r\leq N} |x_1\cdots x_r\rangle$, which, as noted in Section 2.4, is the ground state eigenvector of $\mathcal{H}_r$. The next summand, $S^{(N-1,1)}$, is the $(N-1)$–dimensional subspace generated by functions of the form $A \mapsto \sum_{x\in A} f(x)$, where f is a function on $\{1,\ldots,N\}$ satisfying $\sum_x f(x) = 0$; in physics notation, this is the space generated by an application of $(\sigma^-_{\text{tot}})^{r-1}$ to vectors of the form $\sum_x f(x)|x\rangle \in \mathfrak{H}_1$, where $\sum_x f(x) = 0$, as before. By Proposition 6.5, $E_1 = E_0 + 4\left(\sin\frac{\pi}{N}\right)^2$ is an eigenvalue of $\mathfrak{H}_r$ for all r; since it derives from an eigenvector of $\mathfrak{H}_1$, this eigenvalue is associated to $S^{(N-1,1)}$. Our conjecture, in this language, is that the lowest excited energy eigenvalue of $\mathcal{H}_r$ is associated to $S^{(N-1,1)}$ for any r.

As a side note, another walk which may be of interest is the even more general walk in which more than two colors of cards are allowed; this corresponds to a physical system in which the particles have more than two options for spins. A physical solution to this random walk might rely on Sutherland's Bethe *ansatz* approach towards systems with more than two types of spins [15].

6.3 Applications of random walk analysis to statistical mechanics

We conclude this paper with a promising link, which we believe is new, between mathematical random walks and thermodynamics. The majority of this paper has been devoted to applications of physics, and the Bethe *ansatz* in particular, to random walk theory; in this final section, we reverse the process and demonstrate how random walk theory may be applied to physics, by giving the behavior of a physical system at low temperatures. Although our discussion will focus on the interchange model, it can easily be extended to other Markov chains which also model physical systems.

In random walk theory, much attention is devoted to the sum of powers of eigenvalues of a Markov chain $\sum_{\beta\neq 1} \beta^m$, since this sum is more or less the distance of the chain after m steps from stationarity; see section 4.1. This sum of powers is reminiscent of Boltzmann factors and partition functions from classical thermodynamics, as we now see.

Consider a system of N spin-1/2 particles in a line, with periodic boundary conditions, subject to the Heisenberg ferromagnetic XXX Hamiltonian $\mathcal{H}$ given by (2.5). The time-independent Schrödinger equation postulates that this system will have a well-defined energy equal to one of the energy eigenvalues of $\mathcal{H}$. These energies are not equally probable; qualitatively, the system is more likely to occupy a lower energy state than a higher one. In the absence of outside constraints, the energy levels of $\mathcal{H}$ are occupied according to the Maxwell-Boltzmann distribution of classical statistical mechanics, which states that the probability that the system will occupy energy level E is proportional to $e^{-E/kT}$, where k is the Boltzmann constant and T is the temperature of the system. If the ground state (lowest energy level) of the system is E_0 and the excited states (i.e., all states except for the ground state) are given by $E_1, E_2, \ldots$, then the probability P that the system is in some excited state is

$$P = \frac{\sum_{j>0} e^{-E_j/kT}}{e^{-E_0/kT} + \sum_{j>0} e^{-E_j/kT}} \approx \sum_{j>0} e^{-(E_j - E_0)/kT} = P',$$

where we define P' to be the probability that the system is in some excited state relative to the probability that it is in its ground state, and the approximation, which consists of dropping the sum in the denominator, is valid at small temperatures.

To connect these probabilities to random walks, we use the relation $\beta = 1/2 - E/N$, given by Corollary 4.2, between the eigenvalues β of the interchange walk K_r and the energy levels E of the physical system $\mathcal{H}_r$. Since $\beta_0 = 1$, we obtain

$$P' = \sum_{j>0} e^{-N(1-\beta_j)/kT}.$$

But this is extremely similar to $\sum \beta_j^m$; indeed, the inequalities $\beta \leq e^{\beta - 1} \leq \beta^{1/2}$ when $0.3 \leq \beta \leq 1$, and the fact (see Section 4.3) that all eigenvalues β of K_r are between 0.3 and 1 when $N \geq 12$, show that

$$\sum \beta_j^{N/kT} \leq P' \leq \sum \beta_j^{N/2kT}$$

when $N \geq 12$, so that P' is "sandwiched" between two sums of the form $\sum \beta_j^m$. Thus the probability P that the physical system will be in an excited state at temperature T is approximately the distance from stationarity of the interchange chain after N/kT steps. From our results on the interchange walk, we conclude that, if we assume the system has either one or two opposite spins, then the system is "guaranteed" to be in its ground state (the probability of being in an excited state drops exponentially with T) when kT is of order $1/N^2$.

In this way, results about distances of random walks from stationarity translate into results about probabilities of being in excited states. Given a random walk which is linearly related to a Hamiltonian, as K_r is linearly related to $\mathcal{H}_r$, the amount of time needed for the random walk to reach stationarity translates into the temperature needed for the physical system to be in its ground state, and vice versa. Random walks for which this technique may be used include the Bernoulli-Laplace model of diffusion (which translates directly into a Heisenberg model in which all particles are nearest neighbors) and the Ising model; there are probably many more.

Acknowledgments

I would like to thank my advisor, Prof. Persi Diaconis, for helping to guide my research and for providing consistently insightful suggestions.

References

[1] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Holt, Rinehart and Winston, New York, 1976).

[2] H. Bethe, Theorie der Metalle. Erster Teil. Eigenwerte und Eigenfunktionen der lineären atomischen Kette, *Z. Phys.* **71** (1931), 205-226.

[3] P. Diaconis, *Group Representations in Probability and Statistics* (Institute of Mathematical Statistics, Hayward, 1988).

[4] P. Diaconis and L. Saloff-Coste, Comparison theorems for reversible Markov chains, *Ann. Appl. Prob.* **3** (1993), 696-730.

[5] P. Diaconis and M. Shahshahani, Time to reach stationarity in the Bernoulli-Laplace diffusion model, *SIAM J. Math. Anal.* **18** (1987), 208-218.

[6] P. A. M. Dirac, *The Principles of Quantum Mechanics*, 4rd ed. (Clarendon Press, Oxford, 1958), §58.

[7] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 1, 3rd ed. (Wiley, New York, 1968).

[8] L. Hulthén, Über dar Austauschproblem eines Kristalles, *Ark. Mat. Astron. Fys. B* **26A** (1938), 1-106.

[9] Yu. A. Izyumov and Yu. N. Skryabin, *Statistical Mechanics of Magnetically Ordered Systems* (Consultants Bureau, New York, 1988).

[10] G. D. James, *The Representation Theory of the Symmetric Groups*, Lecture Notes in Mathematics **682** (Springer-Verlag, Berlin, 1978).

[11] F. Keffer, Spin Waves, in *Handbuch der Physik*, ed. H. P. J. Wijn, Band XVIII/2 (Springer-Verlag, Berlin, 1966).

[12] T. M. Liggett, *Interacting Particle Systems* (Springer-Verlag, New York, 1985).

[13] D. C. Mattis, *The Theory of Magnetism*, vol. 1 (Springer-Verlag, Berlin, 1981).

[14] R. K. Pathria, *Statistical Mechanics* (Pergamon Press, Oxford, 1972).

[15] B. Sutherland, Model for a multicomponent quantum system, *Phys. Rev. B* **12** (1975), 3795-3805.

[16] L. A. Takhtadzhan and L. D. Faddeev, The quantum method of the inverse problem and the Heisenberg XYZ model, *Russian Math. Surveys* **34** (1979), No. 5, 11-68.

[17] C. N. Yang, From the Bethe-Hulthén hypothesis to the Yang-Baxter equation, in *The Oskar Klein Memorial Lectures*, vol. 1, ed. G. Ekspong (World Scientific, Singapore, 1991).