Tensor Renormalization Group Meets Computer Assistance

Nikolay Ebel,[a] **Tom Kennedy**[b] **and Slava Rychkov**[a]

[a] *Institut des Hautes Études Scientifiques, 91440 Bures-sur-Yvette, France*
[b] *Department of Mathematics, University of Arizona, Tucson, AZ 85721, USA*

E-mail: ebelnikola@gmail.com, tgk@arizona.edu, slava@ihes.fr

Abstract: Tensor renormalization group, originally devised as a numerical technique, is emerging as a rigorous analytical framework for studying lattice models in statistical physics. Here we introduce a new renormalization map - the 2x1 map - which coarse-grains the lattice anisotropically by a factor of two in one direction followed by a 90-degree rotation. We develop a novel graphical language that translates the action of the 2x1 map into a system of inequalities on tensor components, with rigorous estimates in the Hilbert-Schmidt norm. We define a finite-dimensional "bounding box" called the hat-tensor, and a master function governing its RG flow. Iterating this function numerically, we establish convergence to the high-temperature fixed point for tensors lying within a quantifiable neighborhood. Our main theorem shows that tensors with deviations bounded by 0.02 in 63 orthogonal sectors flow to the fixed point. We also apply the method to specific models - the 2D Ising and XY models - obtaining explicit bounds on their high-temperature phase. This work brings the Tensor RG program closer towards a rigorous, computer-assisted construction of critical fixed points.

Contents

1 Introduction

Tensor renormalization group, born as a numerical tool [1, 2], is becoming a rigorous method to study statistical physics of lattice models [3–5]. The stability of the high-T phase was proved for 2D lattice models in Ref. [3] and for the 3D case in Ref. [5]. Ref. [4] studied the low-T phase of $\mathbb{Z}_2$ symmetric models in 2D. One grand goal of this project, which appears within reach, is a rigorous computer-assisted construction of the nontrivial fixed points describing the critical point of the 2D Ising and other universality classes. This paper will bring us a few steps closer to that goal, as follows:

1. We will introduce a novel tensor RG map which we call the 2x1 map (see Fig. 1), which shrinks the lattice by a factor of 2 in only the horizontal direction, the vertical length being unchanged, after which the lattice is rotated by 90 degrees.[1] The 2x1 map is remarkably simple and as such interesting both numerically and for rigorous studies.

2. We will use the 2x1 map for a new rigorous look at the high-T phase, the problem previously studied in [3] using a more complicated map. Our study here will rely on a new efficient graphical language for describing the structure of rigorous tensor RG maps involving disentangling, like the 2x1 map.

3. Our graphical language allows us to seamlessly translate the definition of the map into a set of bounds on how various parts of the tensor change under the RG map. We can thus easily check that the 2x1 map is a contraction (in a weighted Hilbert-Schmidt norm) in a sufficiently small neighborhood of the high-T fixed point, as is the map from [3].

4. Most importantly, the new graphical language is also ideal for implementing the bounds on a computer. This leads to concrete results about the size of the basin of attraction of the high-T fixed point. The main idea is to use a "bounding box" (called hat-tensor later on). The box has a large but finite number of dimensions corresponding to various "sectors" of the tensor. We define a *master function* $\mathfrak{M}$ which computes how the bounding box varies from one RG step to the next. The master function is iterated numerically until the bounding box starts shrinking. It will then keep shrinking indefinitely, establishing convergence to the high-T fixed point. Using interval arithmetic, such computer-assisted results are fully rigorous.

5. We are thus able to show that *any initial tensor deviating from the high-T fixed point by at most 0.02 in the Hilbert-Schmidt norm in every one of* $N = 63$ *orthogonal sectors that we will define, converges to the high-T fixed point under RG* (Theorem 3.1). While 0.02 may seem like a small number, the full Hilbert-Schmidt norm of the initial deviation can be as large as $0.02\sqrt{N} \approx 0.16$.

6. We also study a couple of specific lattice models (the nearest-neighbor Ising model and the XY model), translating them into tensor network representation. For those models we also obtain rigorous bounds on the size of the high-T phase. See Table 1 for a summary.

[1] The idea that 90 degree rotation is useful for RG maps first appeared in our previous work [6, 7].

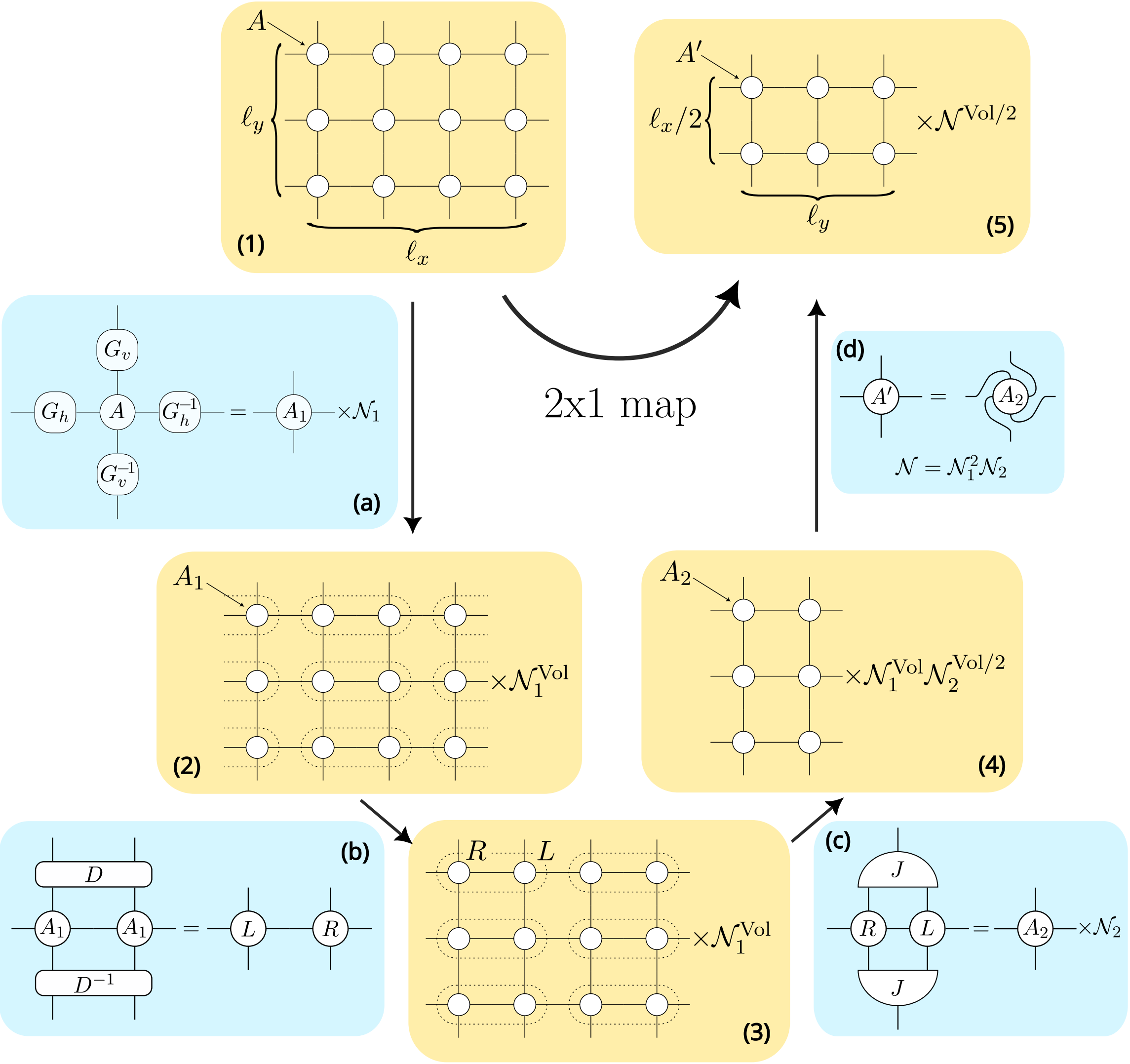


Figure 1: The 2x1 map (see Sec. 2.2 for details) transforms an $\ell_x \times \ell_y$ tensor network made of tensor A into an $\ell_y \times \ell_x/2$ tensor network made of tensor $\mathcal{N}A'$. It is decomposed into 4 steps (blue panels): (a) gauge transformation, (b) disentangling and splitting; (c) reconnection; (d) rotation. The tensor network partition function (assuming periodic boundary conditions) is kept invariant, but is represented differently after each step, as illustrated in yellow panels (1)-(5). Vol $= \ell_x\ell_y$ is the volume of the original lattice.

	2x1 map result	β_c
Ising	$\beta \leqslant 0.12$	$\frac{1}{2}\log(1+\sqrt{2}) \approx 0.44$
XY	$\beta \leqslant 0.18995$	≈ 1.12
General A ($A_{0000}=1$)	$\|A_{abcd}\| \leqslant 0.02$	—

Table 1: High-T phase regions proven here using the 2x1 map for the Ising model, the XY model, and general tensor networks built of a normalized tensor A ($A_{0000} = 1$). For Ising and XY, our bounds may be compared with known β_c (third column). For general tensors, see Section 2.3.1 for the definition of sectors *abcd*.

Our paper builds on the intuition from [3, 4], but it is self-contained. Section 2 is the main technical section which defines the 2x1 map and the master function. After a brief reminder of tensor network notation in Section 2.1, we give an overview of the basic ingredients of the 2x1 map in Section 2.2. Then in Section 2.3 we describe the map at the linearized level, i.e. for small perturbations around the high-T fixed point tensor. In particular in Section 2.3.1 we define "sectors," which are restrictions of tensor to various combinations of subspaces on horizontal and vertical legs. We subdivide tensors into sectors and follow the norm of each sector separately. Section 2.3 culminates in an explanation why the map is a contraction at the linearized level, in an appropriate weighted generalization of the Hilbert-Schmidt norm, and for sufficiently large values of certain parameters $w_{\mathbf{x}}, w_{\mathbf{o}} > 0$ called the reweighting parameters. We then proceed with the nonlinear analysis. In Section 2.4 we define the hat-tensors, which are central for our construction. They play the role of bounding boxes for tensors split into sectors. Hat-tensors reduce the control of the 2x1 map to a finite-dimensional computation of the master function. This basic strategy is explained in Section 2.5. We then proceed to define the 2x1 map at the nonlinear level (Sections 2.6-2.8). This is put together in Section 2.9 where the master function is defined. This is a function from (a subset of) $\mathbb{R}^{63}_{\geqslant 0}$ to itself. In Section 2.10 we come back to the question of how large the reweighting parameters $w_{\mathbf{x}}, w_{\mathbf{o}}$ must be, and derive a necessary condition. We conclude the discussion of the 2x1 map and of its master function in Section 2.11, describing how they behave with respect to the symmetries that the lattice model might have.

In Section 3 we apply the 2x1 map to prove bounds on the high-T phase of the various models in Table 1. The proofs are computer assisted. We provide a short computer code which evaluates the master function. One runs this code for a few iterations, until a condition is verified which implies convergence of subsequent iterations. A short description of the code is given in App. A, and the full documentation is on the GitHub repository [8]. The code follows the paper closely. The equations referenced from the code are given margin tags of the form ⟨code #⟩; see e.g. Eq. (2.1).

To conclude, in Section 4 we discuss various future improvements of the method, which will further enlarge its domain of applicability. There are a few trivial improvements, like optimization of the reweighting parameters $w_{\mathbf{x}}, w_{\mathbf{o}}$. We also describe a strategy for a more dramatic improvement, which will lead to accurate and systematically improvable results for the free energy (in this paper we are content with providing some crude estimates). Finally, we outline how we plan to use a similar improvement of the method to recover the critical point.

2 RG map and the master function

2.1 Tensors

The readers familiar with the tensor notation and terminology (see e.g. [4], Sec. 2) may skip this short section. An n-leg tensor T is a complex multilinear map on $V \times V \times \cdots \times V$. In this paper V will be an infinite dimensional separable complex Hilbert space. So each leg of the tensor corresponds to one of the V's. Equivalently, given an orthonormal basis $\{e_k\}_{k=0}^{\infty}$, we can think of T as a table of complex numbers $T_{i_1,i_2\cdots,i_n} = T(e_{i_1}, e_{i_2}, \ldots, e_{i_n})$. We will sometimes split the legs of the tensor into two groups, say the first k and the last $n-k$, and think of it as a linear operator from $V^{\otimes k}$ to $V^{\otimes n-k}$ acting by $e_{i_1} \otimes \ldots \otimes e_{i_k} \mapsto T_{i_1,i_2\cdots,i_n} e_{i_{k+1}} \otimes \ldots \otimes e_{i_n}$.

The Hilbert-Schmidt (HS) norm of a tensor is $\|T\| = (\sum_{i_1,i_2,\cdots,i_n} |T_{i_1,i_2,\cdots,i_n}|^2)^{1/2}$. Throughout this paper we will encounter three types of tensors:

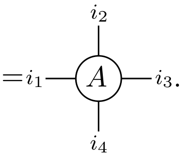
1. Hilbert-Schmidt tensors, i.e. tensors having a finite HS norm;

2. Tensors given by exponentiating a HS tensor with $2n$ legs, viewed as a linear operator on $V^{\otimes n}$. Those will be a sum of an identity operator and of a HS tensor;

3. Isometric tensors which just permute basis elements.

We will study models on a square lattice, and our renormalization group map will act on a HS tensor A with four legs.[2] In the usual graphical notation, we draw it as a blob with four legs representing the tensor indices. The first, second, third, and fourth index will always corresponds to legs pointing left, up, right, and down, respectively, e.g.:

$$A_{i_1 i_2 i_3 i_4} =$$



(2.1)

The graphical notation is convenient to specify tensor contractions, drawn as tensors sharing one or more legs. Such contracted legs are referred to as "bonds". We will also refer to left and right legs as "horizontal" and to up and down legs as "vertical".

2.2 The 2x1 RG map overview

The tensor index 0 will play a special role. We will say that a four-leg tensor is normalized if $A_{0000} = 1$. The 2x1 map will act on a four-leg normalized HS tensor A. It is the RG map in the sense that it preserves the tensor network partition function. The partition function $Z(A, \ell_x \times \ell_y)$ is computed by placing a copy of A at each vertex of the square $\ell_x \times \ell_y$ lattice and contracting legs along the edges, with periodic boundary conditions (see [4], Eq. (2.7)). The result is a complex number, finite if $\ell_x, \ell_y \geqslant 2$ due to A being Hilbert-Schmidt ([4], Prop. 2.3). The 2x1 map will transform A to another tensor $\mathcal{N}A'$, where $\mathcal{N} \in \mathbb{C}$ and A' is normalized, so that the partition function of A on the $\ell_x \times \ell_y$ lattice equals that of $\mathcal{N}A'$ on the $\ell_y \times (\ell_x/2)$ lattice:

$$\boxed{Z(A, \ell_x \times \ell_y) = Z(\mathcal{N}A', \ell_y \times \ell_x/2)}\,. \tag{2.2}$$

(We stress that $\mathcal{N}A'$ is independent of ℓ_x, ℓ_y.) That Z remains the same while the total number of tensors in the network gets reduced is the defining property of any RG map. The unusual feature of the 2x1 map is how the dimensions of the lattice are transformed: $(\ell_x, \ell_y) \mapsto (\ell_y, \ell_x/2)$. This happens because the 2x1 map will first shrink the lattice by a factor of 2 in the horizontal direction, and then rotate by 90 degrees. Applying the 2x1 map twice, the lattice dimensions are both reduced by a factor of 2. One may thus say that the 2x1 map has "effective scale factor $\sqrt{2}$".

Let A_* be the four-leg tensor with only one nonzero component $(A_*)_{0000} = 1$. The infinite-temperature limit of 2D lattice models such as the Ising model corresponds to the tensor network built of A_* [3, 4]. Furthermore, A_* will be a fixed point of the 2x1 RG map. For these reasons we refer to A_* as the high-T fixed point tensor.

We will study the 2x1 map acting on A in a neighborhood of A_*. Since A is normalized, we can write $A = A_* + b$, where $b_{0000} = 0$. Writing also $A' = A_* + b'$, the 2x1 map is thus fully specified by a function $\mathcal{N}(b)$ ("$\mathcal{N}$-factor") and a map $b'(b)$. These will be defined in a neighborhood of 0 of the Hilbert space of complex tensors equipped with the HS norm, and will depend on b analytically.[3] Although physical models often correspond to real tensors, considering complex b is convenient, as it allows to take advantage of this analyticity.

The $\mathcal{N}$-factor trivially factors out of the tensor network partition function. The map $A \mapsto A'$ is called the normalized RG map, and its iterates generate the normalized RG flow. One of our main results ("stability of the high-T fixed point", Theorem 3.1 below) will say that the normalized RG

[2] The letters A and b (see below) will always denote 4-leg tensors. Generic tensors having an arbitrary number of legs will be denoted by T.

[3] See [4], App. A for a few basic facts about analytic functions on Banach spaces.

flow converges to the high-T fixed point for any small initial b. This result is the key to showing that the infinite volume free energy is analytic in the same neighborhood (Proposition 3.3).

The general structure of the 2x1 map is similar but simpler than in [3, 4]. Like the maps in those works, the 2x1 map naturally splits in several steps. The first step (gauge transformation, Fig. 1(a)) multiplies A by G_v and G_v^{-1} on the opposite vertical legs, and by G_h and G_h^{-1} on the opposite horizontal legs. Here G_v and G_h are bounded linear operators with bounded inverses. The partition function is preserved for any G_v, G_h ([4], Sec.2.4.3). We will use G_v and G_h depending on b in a way specified in Sec. 2.6. This step produces tensor $A_g = \mathcal{N}_1 A_1$ with A_1 normalized.

The second step (disentangling and splitting, Fig. 1(b), Sec. 2.7) groups A_1 tensors into horizontally contracted pairs, multiplies the upper vertical legs of the pair by a "disentangler" map $D : V \otimes V \to V \otimes V$, and multiplies the lower vertical legs by its inverse D^{-1}. The resulting 6-leg tensor is then split into a contraction of two 4-leg HS tensors L and R. The partition function is preserved ([4], Sec.2.4.4). All these tensors will be expressed as power series in $b_1 = A_1 - A_*$.

The third step (reconnection, Fig. 1(c), Sec. 2.8) contracts tensors R and L horizontally, in the opposite order from how these tensors were obtained during disentangling. The upper vertical legs are then contracted with an appropriately chosen isometry $J : V \otimes V \to V$, and the lower vertical legs with its transpose. The resulting tensor is denoted $\mathcal{N}_2 A_2$ with A_2 normalized. Since $J^T J = \mathbb{1}$, the partition function is preserved. This is the step which halves the horizontal lattice size .

The final step (rotation, Fig. 1(d), Sec. 2.8) defines the tensor A' as the counterclockwise rotation of A_2 by 90 degrees. This completes the overview of the 2x1 map. The $\mathcal{N}$-factor is given by $\mathcal{N} = \mathcal{N}_1^2 \mathcal{N}_2$.

2.3 Linearized analysis

We start in this section by considering the linearized 2x1 map (i.e. to first order in b). The map was designed so that this linearization is a contraction (with respect to a norm defined below). We would like that the reader be convinced about that, before confronting the somewhat technical discussion of the full nonlinear map in the subsequent sections.

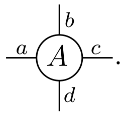
2.3.1 Sectors

We will decompose the tensor leg Hilbert spaces into a direct sum of orthogonal subspaces called "sectors". This is needed both at the linear and, later, at the nonlinear level. On the vertical legs, we decompose $V = \mathbf{o} \oplus \mathbf{x}$ where $\mathbf{o}$ is the one-dimensional sector spanned by e_0, and $\mathbf{x}$ is the infinite-dimensional orthogonal complement of $\mathbf{o}$, spanned by $\{e_k\}_{k=1}^{\infty}$. On the horizontal legs we decompose $V = \mathbf{o} \oplus \mathbf{d} \oplus \mathbf{u} \oplus \mathbf{r}$ where $\mathbf{o}$ is as before, and $\mathbf{d}$, $\mathbf{u}$, $\mathbf{r}$ are three infinite-dimensional sectors. (Their names stand for down, up, rest.)[4] Why we use two vertical sectors and four horizontal sectors will become clear in the course of our analysis, see in particular Section 2.3.4. We will sometimes use $\mathbf{x} = \mathbf{d} \oplus \mathbf{u} \oplus \mathbf{r}$ on the horizontal legs.

Furthermore, we will decompose the tensors A on which the 2x1 map acts into a collection of tensors obtained by restricting tensor legs to the above-defined sectors. Namely, for a four-leg tensor A, given $a, c \in \{\mathbf{o}, \mathbf{x}\}$ and $b, d \in \{\mathbf{o}, \mathbf{d}, \mathbf{u}, \mathbf{r}\}$, we define the restricted tensor A_{abcd} as the tensor whose components are equal to those of A when all four indices are in the respective sectors and equal zero otherwise. These restricted tensors will also be referred to as "sectors" of the original tensor. So there are $4 \times 2 \times 4 \times 2 = 64$ sectors for A. Obviously, the sum of A_{abcd} over all 64 possibilities for $abcd$ equals A. It will be sometimes convenient to set $b, d = \mathbf{x}$ on the horizontal legs. This will correspond to the sum of sectors over $b, d \in \{\mathbf{d}, \mathbf{u}, \mathbf{r}\}$. In graphical notation A_{abcd} will be denoted by putting sector labels *next* to the corresponding legs (unlike indices of tensor elements in (2.1),

[4] The exact numbering of basis vectors which span $\mathbf{d}, \mathbf{u}, \mathbf{r}$ will not be important. E.g. one can choose $\mathbf{d} = \mathrm{span}(\{e_{3i+1}\}_{i=0}^{\infty})$, $\mathbf{u} = \mathrm{span}(\{e_{3i+2}\}_{i=0}^{\infty})$, $\mathbf{r} = \mathrm{span}(\{e_{3i+3}\}_{i=0}^{\infty})$.

put *at the extremities* of the legs):

$A_{abcd} =$  . (2.3)

This definition extends to general tensors. If T is an n-leg tensor and $a_1, a_2, \cdots, a_n$ are sectors for each of the legs, then the restricted tensor $T_{a_1,a_2,\cdots,a_n}$ is the tensor whose components are equal to those of T when all of the indices are in the respective sectors and equal zero otherwise.

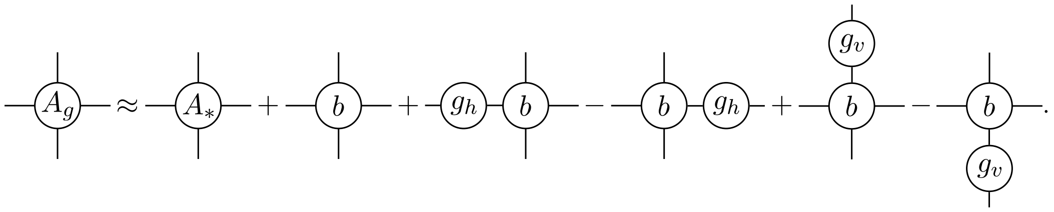
2.3.2 Gauge transformation

We now proceed with the linearized analysis of the 2x1 map. We write $A = A_* + b$, $A' = A_* + b'$. Our purpose is to compute b' to first order in b.

The first step of the 2x1 map (Fig. 1(a)) is a gauge transformation. Since our final goal is to reduce b, let us try to somewhat reduce it via the gauge transformation before doing something more sophisticated. We choose the gauge transformation matrices G_h and G_v so that[5]

$$G_h \approx \mathbb{1}_V + g_h, \qquad G_v \approx \mathbb{1}_V + g_v\,, \tag{2.4}$$

with g_h, g_v linear in b. We apply G_h, G_v and their inverses to the horizontal, vertical legs of A, respectively, as shown in Fig. 1(a). We denote the resulting tensor by A_g. It is easy to see that

 (2.5)

We define the two-leg tensors g_h and g_v by the following expressions:

$$g_h = b_{\mathbf{o}0\mathbf{x}0} - b_{\mathbf{x}0\mathbf{o}0}, \quad g_v = b_{0\mathbf{o}0\mathbf{x}} - b_{0\mathbf{x}0\mathbf{o}}. \tag{2.6}$$

A word about the notation here. The tensor b has four legs. We can form two-leg tensors from it by setting the indices on two of the legs to 0. For example, $b_{\mathbf{x}0\mathbf{o}0}$ means that the vertical legs are fixed to be 0, the left leg is restricted to the $\mathbf{x}$ sector[6] and the right leg is restricted to the $\mathbf{o}$ sector. (We recall that restricting the leg to the $\mathbf{o}$ sector is not the same as setting the index of the leg to 0. The former leaves the leg, the latter removes it.) $b_{\mathbf{o}0\mathbf{x}0}$ is similar, except that now the left leg is restricted to the $\mathbf{o}$ sector and the right leg to the $\mathbf{x}$ sector. So $b_{\mathbf{x}0\mathbf{o}0}$ and $b_{\mathbf{o}0\mathbf{x}0}$ are tensors with only horizontal legs. Similar definitions for $b_{0\mathbf{x}0\mathbf{o}}$ and $b_{0\mathbf{o}0\mathbf{x}}$ give tensors with only vertical legs. Eq. (2.6) defines g_h and g_v as linear combinations of these tensors. Now, it's easy to see that with this definition the first four diagrams in (2.5) cancel precisely the sectors

$$b_{\mathbf{xooo}}, b_{\mathbf{oxoo}}, b_{\mathbf{ooxo}}, b_{\mathbf{ooox}} \tag{2.7}$$

of the last diagram, while the other sectors are left unchanged (at first order). This accomplishes the stated goal of "somewhat reducing b".[7] We also have $\mathcal{N}_1 = (A_g)_{0000} \approx 1$. So the normalized tensor $A_1 = A_g/\mathcal{N}_1$ is written as $A_* + b_1$ where $b_1 \approx b$ with sectors (2.7) set to zero. When we describe the subsequent steps, we simply rename A_1 and b_1 as A and b and assume that the sectors (2.7) are zero.

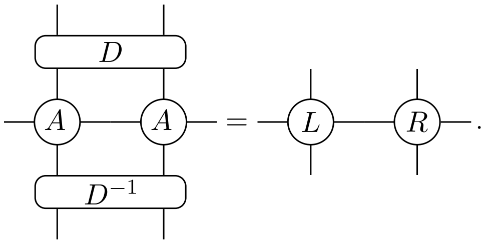
[5] In this section, $\approx$ means modulo b^2 terms.

[6] Here we are using the convention that $\mathbf{x}$ on horizontal legs stands for $\mathbf{d} \oplus \mathbf{u} \oplus \mathbf{r}$.

[7] This is the best (at first order in b) one can do with a gauge transformation.

2.3.3 Disentangling

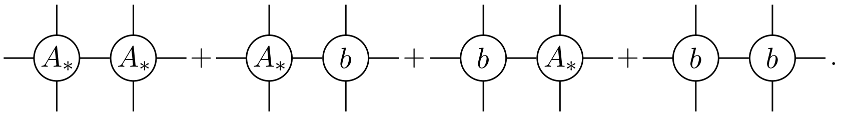
The second step of our RG map is disentangling and splitting, shown in Fig. 1(b) which we copy here (recall that we renamed A_1 as A and b_1 as b):



(2.8)

Our goal will be to compute L and R to first order in b. For this we will need to include some diagrams which are second and third order in b. This is different from the previous section on gauge transformation where all considered diagrams were first order in b.

When we contract the two A tensors in (2.8) and expand in powers of b, we get four diagrams:

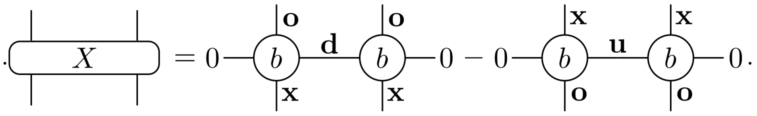


(2.9)

The disentangler will be chosen as

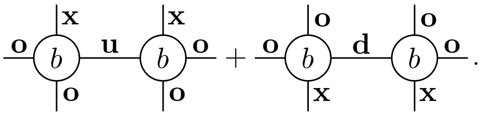
$$D = \mathbb{1}_{V\otimes V} + X + O(b^4), \tag{2.10}$$

where $X = O(b^2)$ is given by



(2.11)

When we act with D and with $D^{-1} = \mathbb{1} - X + O(b^4)$ on the contraction of two A's, the important term to the considered order in b is when X is contracted to the first diagram in (2.9). This term cancels the following part of the last diagram in (2.9):



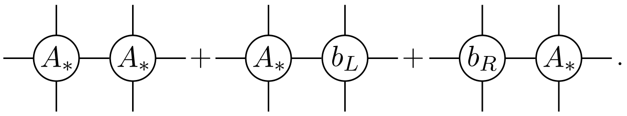
(2.12)

The 4 sectors of b involved here have one horizontal leg **u** or **d**, one vertical leg **x**, and two other legs **o**. Furthermore, the position of the **x** leg (up or down) is correlated with whether it's **u** or **d** on the horizontal leg. This correlation is the raison d'être of the **u** and **d** sectors. We will see below how it is generated naturally during the reconnection and rotation steps of the map.

We still have to perform the splitting, i.e. find L and R so that Eq. (2.8) holds. We write $L = A_* + b_L$, $R = A_* + b_R$, and we have to find b_L, b_R to linear order in b. Let us postpone this task and see first what happens when we perform the reconnection and rotation.

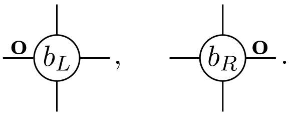
2.3.4 Reconnection and rotation

Reconnection, Fig. 1(c), contracts L and R in the opposite order and acts with an isometry J on the pairs of horizontal legs. At the linearized level the contraction amounts to the three diagrams:



(2.13)

The only important sectors of b_L, b_R are those which contribute to this contraction, which are



(2.14)

Let us discuss the isometry J. It's very natural to choose this isometry so that it maps sector $\mathbf{o}\otimes\mathbf{o}$ to $\mathbf{o}$. With this choice, the first diagram in (2.13) is mapped to A_*.

The other two diagrams in (2.14) have an interesting structure: the former has **ox** or **oo** on the vertical pairs of legs (but not both **oo** for the reason stated immediately after), while the latter satisfies a similar rule with **ox** replaced by **xo**. "Both **oo**" are excluded because, see Section 2.3.6,

$$(b_L)_{\mathbf{oooo}}, (b_L)_{\mathbf{ooxo}}, (b_R)_{\mathbf{oooo}}, (b_R)_{\mathbf{xooo}} \approx 0\,. \tag{2.15}$$

To take advantage of the described structure, we choose the isometry J so that[8]

$$\mathbf{o}\otimes\mathbf{x}\to\mathbf{u}, \qquad \mathbf{x}\otimes\mathbf{o}\to\mathbf{d}\,. \tag{2.16}$$

Then, after counterclockwise 90 degree rotation, as in Fig. 1(d), we will obtain diagrams in the following 8 sectors, referred below as "special":

$$\mathbf{oxuo}, \mathbf{uxoo}, \mathbf{oodx}, \mathbf{doox}, \mathbf{uxuo}, \mathbf{uouo}, \mathbf{dodx}, \mathbf{dodo}\,. \tag{2.17}$$

The first four special sectors are the ones involved in the canceled diagram (2.12).

2.3.5 When is the linearized map a contraction?

We just showed that after the RG step b' has only the special components (2.17) nonzero (to first order). This leads to some simplification of the analysis when the linearized RG map is a contraction, and with respect to which norm. Let P_s (P_n) be the orthogonal projectors on the sum of special (non-special) sectors. Let ∇R be the Jacobian of the RG map at A_*. We decompose $b = b_s + b_n$ where $b_s = P_s b$, $b_n = P_n$. The important part of the Jacobian is $K_{ss} = P_s(\nabla R)P_s$, which shows how b'_s depends on b_s (to first order). The less important part is $K_{sn} = P_s(\nabla R)P_n$ which shows how b'_s depends on b_n. The K_{ss} and K_{sn} are some bounded linear operators (with respect to the HS norm). Consider the following weighted norm:

$$|||b|||= \max(\|b_s\|, C\|b_n\|), \tag{2.18}$$

where $\|\cdot\|$ is the HS norm and $C > 0$ is a constant to be determined. We have

$$|||(\nabla R)b|||= \|((\nabla R)b)_s\| \leqslant \|K_{ss}\|\|b_s\| + \|K_{sn}\|\|b_n\| \leqslant \nu|||b|||, \quad \nu = \|K_{ss}\| + C^{-1}\|K_{sn}\|\,. \tag{2.19}$$

We see that if $\|K_{ss}\| < 1$ and $\|K_{sn}\|$ is finite, we can always choose C so that $\nu < 1$ and ∇R is a contraction with respect to $|||\cdot|||$. That we only need to worry about $\|K_{ss}\| < 1$ is the above-mentioned simplification.

2.3.6 b_L, b_R and the norm of K_{ss}

We come back to the task of defining $L = A_* + b_L$ and $R = A_* + b_R$. We will also see how the condition $\|K_{ss}\| < 1$ comes about.

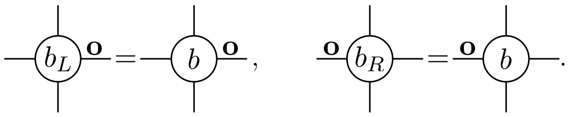
The idea to define L and R is simple: we have to "cut" the l.h.s. of Eq. (2.8) vertically into two halves, assigning the left half to L, and the right half to R. The contraction of two A's and the

[8] The fourth combination $\mathbf{x}\otimes\mathbf{x}$ does not occur in the linear analysis. In the nonlinear analysis it will be mapped by J to the sector $\mathbf{r}$ introduced specifically for this purpose.

diagrams defining X can both naturally be cut in two halves. We also have to take into account that some diagrams cancel. We expand in powers of b. Let n_L (resp. n_R) denote the order of the left (resp. right) half of a diagram in b. Since here we only need b_L and b_R to first order, we have to keep diagrams where $n_L \leqslant 1$ or $n_R \leqslant 1$.

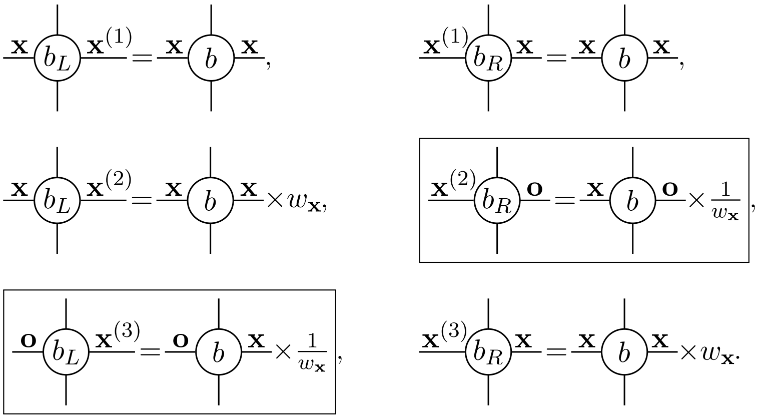
We start by discussing the diagrams obtained by expanding Eq. (2.8) to $O(b^2)$. These are the diagrams in (2.9), with diagrams from (2.12) cancelled out. These are the most important diagrams. A few remaining diagrams, which are $O(b^3)$, will be dealt with below.

Consider first the sum of the first three diagrams in (2.9), and of the fourth diagram with the contracted leg restricted to $\mathbf{o}$. We refer to this sum as the "z channel" (z for zero). This channel is reproduced by defining



(2.20)

Consider next the fourth diagram in (2.9) with the contracted leg restricted to $\mathbf{x}$. There are four cases ("channels") to consider, the external horizontal legs being $\mathbf{xx}$, $\mathbf{xo}$, $\mathbf{ox}$, $\mathbf{oo}$. The first three channels are reproduced by defining (see below for the meaning of boxes around some equations)

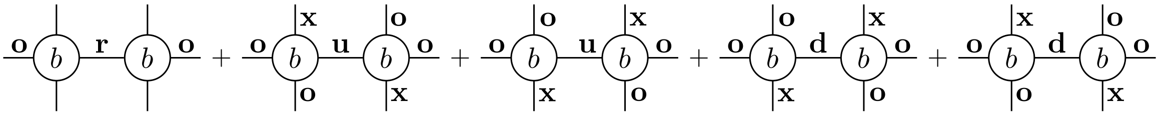


x b_L $\mathbf{x}^{(1)}$ = x b x ,
$\mathbf{x}^{(1)}$ b_R x = x b x ,
x b_L $\mathbf{x}^{(2)}$ = x b x $\times w_{\mathbf{x}}$,
$\mathbf{x}^{(2)}$ b_R o = x b o $\times \frac{1}{w_{\mathbf{x}}}$,
o b_L $\mathbf{x}^{(3)}$ = o b x $\times \frac{1}{w_{\mathbf{x}}}$,
$\mathbf{x}^{(3)}$ b_R x = x b x $\times w_{\mathbf{x}}$.

(2.21)

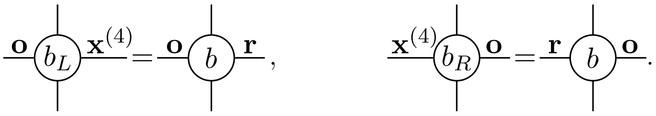
Here we introduced a new notation $\mathbf{x}^{(i)}$ which denotes a unique copy number i of the $\mathbf{x}$ sector.[9] Copies need to be created to avoid unwanted cross terms which would otherwise arise between b_L and b_R diagrams for different channels. We also introduced a reweighting parameter $w_{\mathbf{x}} > 0$ for the $\mathbf{xo}$ and $\mathbf{ox}$ channels. Playing with this parameter will help us achieve $\|K_{ss}\| < 1$ (see below).

Finally consider the $\mathbf{oo}$ channels. In this sector we have cancellations: the diagrams (2.12) have been canceled by the disentangler. We are left with the five diagrams



(2.22)

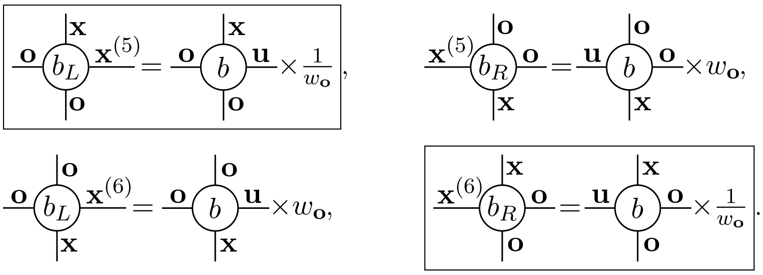
(Note that we omitted some diagrams which are higher order because $b_{\mathbf{x}000} = b_{00\mathbf{x}0} = 0$, to first order, as a result of the gauge transformation.) The first three diagrams are reproduced by setting (note that we are introducing more sector copies):



o b_L $\mathbf{x}^{(4)}$ = o b r ,
$\mathbf{x}^{(4)}$ b_R o = r b o .

(2.23)

[9] More formally, a copy of $\mathbf{x}$ is a tensor product of $\mathbf{x}$ with a one-dimensional auxiliary Hilbert space.



(2.24)

The last two diagrams in (2.22) are reproduced analogously to Eq. (2.24) by defining more copies $\mathbf{x}^{(7)}, \mathbf{x}^{(8)}$; we omit the corresponding equations. In (2.24), we introduced another reweighting parameter $w_\mathbf{o} > 0$, needed for $\|K_{ss}\| < 1$.

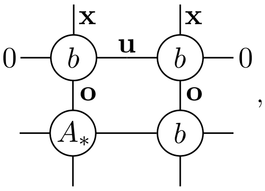
So let us see how $\|K_{ss}\| < 1$ is obtained given the above assignments of b_L and b_R sectors, and with the help of reweighting parameters. We are only interested in the sectors of b_L and b_R that satisfy both of the following two conditions:

- They are the "important" sectors from Eq. (2.14), i.e. have $\mathbf{o}$ on the "external" horizontal leg. Only these sectors contribute to b' at linear order.
- They are proportional to b from the special sectors. Only these sectors contribute to K_{ss}.

There are only four b_L and b_R sectors in Eqs. (2.20),(2.21),(2.23), (2.24) which satisfy both of these conditions; we put boxes around the equations defining them. The key point is that all these sectors are rescaled by $1/w_\mathbf{x}$ or $1/w_\mathbf{o}$. Choosing $w_\mathbf{x}$ and $w_\mathbf{o}$ large we will suppress them and achieve $\|K_{ss}\| < 1$. The remaining sectors do not matter since they either do not contribute to b' at linear order, or contribute only to K_{sn} and not to K_{ss}.

This is also a good moment to see why we needed to disentangle. If we did not cancel the diagrams (2.12) by the disentangler, we would need additional b_L, b_R sectors to reproduces them. But both halves of those diagrams belong to special sectors, hence they both contribute to K_{ss}. Therefore, there would be no way to achieve $\|K_{ss}\| < 1$ by reweighting.[10]

This concludes the treatment of the diagrams up to $O(b^2)$. Consider next the following diagram



(2.25)

which appears when we expand the l.h.s. of Eq. (2.8) at $O(b^3)$. It has $n_L = 1$, $n_R = 2$, so when we cut it vertically, it contributes to b_L at linear order. The left leg of A_* being in $\mathbf{o}$, this b_L will contribute to b'. To suppress it, we use reweighting, which works here since b_R is second order. For simplicity let us use the same reweighting parameters: $w_\mathbf{o}$ if the right external leg is also in $\mathbf{o}$, and $w_\mathbf{x}$ if it's in $\mathbf{x}$. There are 3 more $O(b^3)$ diagram like (2.25), all treated in the same way.

We still owe a check that the b_L, b_R sectors mentioned in (2.15) vanish to first order. This follows by the inspection of (2.20),(2.21),(2.23),(2.24), and the diagrams obtained by cutting the cubic terms like (2.25). There are only a few diagrams which contribute to (2.15), and they are all proportional to the corresponding sectors of b. These sectors are either zero ($b_{\mathbf{oooo}}$), or have been set to zero to first order by the gauge transformation.

This finishes the discussion of b_L, b_R up to linear order. As explained we can make $\|K_{ss}\| < 1$ by choosing $w_\mathbf{x}$, $w_\mathbf{o}$ sufficiently large. An explicit condition on $w_\mathbf{x}$, $w_\mathbf{o}$ for this to happen will be

[10] A related comment is that tensor RG maps without disentangling ("simple RG maps") typically have eigenvalue 1 perturbations at the high-T fixed point, associated with the so called CDL tensors; see [3].

worked out in Section 2.10 below. We then choose C sufficiently large so that $\nu < 1$ in (2.19) and the linearized RG map is a contraction with respect to the $|||\cdot|||$ norm.

Hopefully, it is now less mysterious how we designed our RG map and why we think it's a good map. We will now proceed to the discussion of the map at the full nonlinear level. This will necessarily be more complicated, but the main ideas will be the same. There, we will also introduce a graphical language which will automatize the task of keeping track of various diagrams and making sure that no diagram is forgotten.

2.4 Hat-tensors

As we have already seen in the linearized discussion, various tensors will be separated into sectors. An essential ingredient of this paper will be to control the HS norm of each sector separately. (We have already seen a simple example of this in the weighted norm (2.18).) This leads to more refined estimates than using the overall HS norm of b as in [3–5].

Definition 2.1. *Let T be a HS tensor with n legs. We will say that an n-leg tensor $\widehat{T}$ is a hat-tensor for T if*

$$\|T_{a_1 a_2 \cdots a_n}\| \leqslant \widehat{T}_{a_1 a_2 \cdots a_n} \,, \tag{2.26}$$

where each index a_i ranges over the sectors for the index space for that leg. (So $\widehat{T}$ is a finite dimensional tensor with real nonnegative components.)

Remark 2.1. A hat-tensor always exists for a HS tensor since we can simply define $\widehat{T}_{a_1,a_2,\cdots,a_n}$ to be $\|T_{a_1,a_2,\cdots,a_n}\|$. We call this hat-tensor the minimal hat-tensor for T.

We now have two types of tensors: the original tensors for which V is infinite dimensional and the hat-tensors for which the index spaces are finite dimensional. When we want to emphasize that we are referring to the former, we will call them "full-tensors." We will label indices for the full-tensors with letters such as $i_1, i_2, \ldots$. For the hat-tensors the possible values of the indices are the sectors for the corresponding leg. We will label indices for the hat-tensors with letters such as $a, b, c, \ldots$ or $a_1, a_2, \ldots$.

It is immediate that if $\widehat{A}$ and $\widehat{B}$ are hat-tensors for A and B, then $\widehat{A} + \widehat{B}$ is a hat-tensor for $A + B$. We need to consider hat-tensors for contractions of tensors. Note that we will only contract legs of different tensors (perhaps even several legs simultaneously), never two legs of the same tensor ("no self-contraction").

Lemma 2.1. *Let $T_1, T_2, \ldots, T_n$ be HS tensors and let C be some contraction of these tensors without self-contractions. Let $\widehat{T}_i$ be hat-tensors for T_i. Then the contraction of $\widehat{T}_1, \widehat{T}_2, \ldots, \widehat{T}_n$ is a hat-tensor for C. (The $\widehat{T}_i$ are contracted in the same way as the T_i.)*

Sketch of the proof: Proposition 2.4 in [4] is a partial case of this lemma with one sector per leg, when the hat-tensor is just a scalar bound for the full HS norm. Like that proposition, the lemma is a simple consequence of the Cauchy-Schwarz inequality. A simple induction argument reduces the proof to the case of $n = 2$. For this case we group together the legs of A_1 that are not involved in the contraction, and group together the legs of A_1 that are involved. We do a similar leg grouping for A_2. This reduces the $n = 2$ case to the case of two 2-leg tensors. This case can then be proved by several applications of the Cauchy-Schwarz inequality, paying attention to different sectors. □

As mentioned in Section 2.1, we will also have some tensors which are not HS and so do not have hat-tensors. To handle these we make the following definition.

Definition 2.2. *Let T be a tensor (not necessarily HS) with some non-empty subset of its legs distinguished. We say $\check{T}$ is a check-tensor for T with this subset of distinguished legs if for any HS* *tensor S with hat-tensor $\widehat{S}$, the contraction of the distinguished legs of T with any subset of the legs of S produces a HS tensor, and a hat-tensor for this contraction may be obtained by contracting $\check{T}$ and $\widehat{T}$ in the same manner.*

Remark 2.2. In our diagrams we will put a tick mark across the distinguished legs for tensors that have a check-tensor.

Remark 2.3. The idea behind this definition that T is a bounded operator from the (Hilbert space of) distinguished legs to the undistinguished legs, and $\check{T}$ bounds the operator norm of restrictions of T to various sectors. This generalizes Prop. 2.2(b) from [4].

Using Lemma 2.1 and Definition 2.2 repeatedly we can compute hat-tensors for a wide class of *allowed* contractions defined as follows:

Definition 2.3. *Let $n \geqslant 1$ and let $T_1, T_2, \dots, T_n$ be HS tensors with hat-tensors $\widehat{T}_i$. Let $m \geqslant 0$ and let $S_1, S_2, \dots S_m$ be tensors each of which is assigned a set of distinguished legs and has a check-tensor $\check{S}_j$ for that set. A contraction C of all these tensors is called* allowed *if there are no self-contractions, and for each S_j the set of distinguished legs is contracted with some subset of the legs of a single tensor T_i, while the undistinguished legs of S_j may be either contracted with legs of tensors $T_{i'}$, $i' \neq i$, or left uncontracted.*

The following lemma will be used extensively.

Lemma 2.2. *Let C be an allowed contracted from the previous definition. Then C is HS and if we contract the $\widehat{T}_i$ and the $\check{S}_j$ in the same manner as the contraction that formed C, then we obtain a hat-tensor for C.*

Proof. Let i be such that the distinguished legs of S_1 are contracted with some legs of T_i. Then by the definition of check-tensor, we may replace the contraction between S_1 and T_i by a single HS tensor for which the contraction of $\check{S}_1$ and $\widehat{T}_i$ is a hat-tensor. This reduces m by 1 and leaves n unchanged. Continuing this process we are reduced to the case of $m = 0$. Then we can apply the previous lemma. □

Remark 2.4. Whether one uses a hat-tensors or a check-tensors is sometimes a matter of necessity and sometimes of convenience. In particular, in this paper:

- We use a hat-tensor for any tensor which is a contraction of b tensors. This hat-tensor will be obtained from $\widehat{b}$ by Lemma 2.1.

- We use a check-tensor for any two-leg tensor T which is an orthogonal projector on the direct sum of one or more sectors. For example, if T is the projector on $\mathbf{d} \oplus \mathbf{u} \oplus \mathbf{r}$, then its check-tensor $\check{T}_{aa} = 1$, $a = \mathbf{d}, \mathbf{u}, \mathbf{r}$, and all other components equal to zero. We find it convenient to use a check-tensor even for projectors on finite-dimensional subspaces. even though these tensors are HS and a hat-tensor exists. For example, if T is the projector on $\mathbf{o}$ then $\check{T}_{\mathbf{oo}} = 1$ is the only nonzero component.

2.5 Control of the 2x1 map via hat-tensors

Let us describe the main ideas of how we will control the 2x1 RG map on the nonlinear level. We denote by $\mathbb{H}_0$ the Hilbert space of complex tensors b equipped with the HS norm and satisfying $b_{0000} = 0$. The RG map acts on the tensor $A = A_* + b$, where $b \in \mathbb{H}_0$.

Let $\widehat{b}$ be a hat-tensor for b (which may or may not be its minimal hat-tensor). It is a tensor with $2 \times 4 \times 2 \times 4 = 64$ real nonnegative components, one of which is zero: $b_{\mathbf{oooo}} = 0$. We denote the set of such hat-tensors $\widehat{\mathbb{H}}_0$. We will perform the four steps of the RG map one by one, and after every step we will estimate the arising new tensors by their hat-tensors, which will be expressed as functions of previous hat-tensors:

- After the gauge transformation step we will obtain $A_1 = A_* + b_1$ with a hat-tensor $\widehat{b}_1$ which is a function of $\widehat{b}$;
- After disentangling and splitting we'll get $L = A_* + b_L$, $R = b_R$ with hat-tensors $\widehat{b}_L$, $\widehat{b}_R$ which are functions of $\widehat{b}_1$;
- After reconnection we'll get $A_2 = A_* + b_2$ with a hat-tensor $\widehat{b}_2$ a function of $\widehat{b}_L$ and $\widehat{b}_R$;
- And finally after rotation we have $A' = A_* + b'$ with $\widehat{b}'$ a function of $\widehat{b}_2$.

All these functions will be given by explicit expressions which will be easily inferred from the diagrams defining the RG map, given below, using Lemma 2.2. Composing the functions from the four steps of the 2x1 map, we get $\widehat{b'}$ as a function of $\widehat{b}$, which we call the master function:

$$\widehat{b'} = \mathfrak{M}(\widehat{b})\,. \tag{2.27}$$

The master function acts from $\widehat{\mathbb{H}}_0$ to itself. Because of the large dimensionality of this space, there is no question of computing it by hand. We realize it in a computer code accompanying the paper.

Our construction will define the 2x1 map for b belonging to an open neighborhood Ω of $0 \in \mathbb{H}_0$. Moreover b' will be an analytic function of $b \in \Omega$. The neighborhood Ω will be defined in terms of a condition on a hat-tensor for b:

$$\Omega = \{b : \exists\ \widehat{b} \text{ for } b \text{ such that } \widehat{b} \in \widehat{\Omega}\}\,. \tag{2.28}$$

where $\widehat{\Omega}$ is a subset of $\widehat{\mathbb{H}}_0$. It will be easy to verify condition $\widehat{b} \in \widehat{\Omega}$ for any numerical tensor $\widehat{b}$.[11]

Now suppose we are given a tensor b whose hat-tensor $\widehat{b}$, specified numerically, belongs to $\widehat{\Omega}$, so the 2x1 map is defined. We apply the map and we get b'. The b' will be defined explicitly by some diagrams in terms of b. Of course b', like b, is an infinite-dimensional tensor, so we don't attempt to evaluate those diagrams. Instead we evaluate the master function and obtain a numerical hat-tensor $\widehat{b'}$, which gives us information about the size of b'. In particular we can check if $\widehat{b'} \in \widehat{\Omega}$. If this condition holds, this implies that $b' \in \Omega$, and we can iterate the 2x1 map. We so obtain a sequence of tensors $b = b^{(0)}, b' = b^{(1)}, b^{(2)}, b^{(3)}, \ldots$ (which are just defined, not evaluated), and the corresponding hat-tensors, which are all evaluated numerically.

We will find that when we iterate the map one of the following two possibilities is realized:

After a certain number of iterations we get a hat-tensor $\widehat{b}^{(i_0)} \notin \widehat{\Omega}$. (2.29)

The iterations enter a trajectory where all components $\widehat{b}^{(i)}$ start decreasing exponentially with i, in the sense that for some i_0 the condition $\widehat{b}^{(i_0+1)} \leqslant \lambda \widehat{b}^{(i_0)}$ is satisfied componentwise with some $\lambda < 1$. (2.30)

In the first case we have to stop the iterations as the RG map is no longer defined. In the second case, as we will argue now, the exponential decrease will continue indefinitely.

The argument is based on some properties of functions of hat-tensors. We need a few definitions. If $x, y \in \mathbb{R}^n_{\geqslant 0}$ (e.g. they are both hat-tensors), then we say $x \leqslant y$ if this inequality holds componentwise. Used subsets of $\mathbb{R}^n_{\geqslant 0}$ will satisfy the following definition:

Definition 2.4. *A set $\widehat{\Lambda} \subset \mathbb{R}^n_{\geqslant 0}$ is called* **downward closed** *if for any $x \in \widehat{\Lambda}$ and for any y such that $0 \leqslant y \leqslant x$ we have $y \in \widehat{\Lambda}$.*

The following properties will hold naturally for the functions of hat-tensors that we will consider.

[11] The $\widehat{\Omega}$ will be downward closed (Def. 2.4 below). So to check if $b \in \Omega$ it's enough to check if its minimal hat-tensor belongs to $\widehat{\Omega}$. Still, it's convenient to write the definition for Ω as above.

Definition 2.5. *Let $\widehat{\Lambda} \subset \mathbb{R}^n_{\geqslant 0}$ be downward closed. A map $f : \widehat{\Lambda} \to \mathbb{R}^m_{\geqslant 0}$ is called* **monotonic** *if the inequality $f(y) \leqslant f(x)$ holds componentwise for all $y, x \in \widehat{\Lambda}$ such that $y \leqslant x$ componentwise. The map is called* **subhomogeneous** *if the inequality $f(\mu x) \leqslant \mu f(x)$ holds componentwise for all $x \in \widehat{\Lambda}$ and all $0 \leqslant \mu \leqslant 1$.*

In the course of Sections 2.6 to 2.9, we will present explicit definitions of $\mathfrak{M}$ and $\widehat{\Omega}$, together with a proof that $\mathfrak{M}$ is monotonic and subhomogeneous on $\widehat{\Omega}$. The following key lemma then implies the convergence of $\widehat{b}^{(i)}$ satisfying (2.30).

Lemma 2.3 (Key lemma)**.** *Let $\widehat{b}^{(i)}$ be the sequence of hat-tensors generated by iterating the master function, assumed monotonic and subhomogeneous. Suppose (2.30) is realized. Then for all $i > i_0$,*

$$\widehat{b}^{(i)} \leqslant \lambda^{i-i_0} \widehat{b}^{(i_0)}, \tag{2.31}$$

and thus $\widehat{b}^{(i)}$ converges to zero exponentially fast.

Proof. We proceed by induction on i. For $i = i_0 + 1$ the inequality holds by (2.30). The induction step $i \to i+1$ is a consequence of monotonicity and subhomogeneity:

$$\begin{aligned} \widehat{b}^{(i+1)} &= \mathfrak{M}(\widehat{b}^{(i)}) \leqslant \mathfrak{M}(\lambda^{i-i_0}\widehat{b}^{(i_0)}) && \text{(monotonicity)} \qquad (2.32)\\ &\leqslant \lambda^{i-i_0}\mathfrak{M}(\widehat{b}^{(i_0)}) && \text{(subhomogeneity)} \qquad (2.33)\\ &\equiv \lambda^{i-i_0}\widehat{b}^{(i_0+1)} \leqslant \lambda^{i+1-i_0}\widehat{b}^{(i_0)}, && \qquad (2.34) \end{aligned}$$

completing the induction. □

The following simple facts about monotonic, subhomogeneous maps will be useful below.

Lemma 2.4. *(a) Any monomial of degree $\geqslant 1$ in components of $\widehat{b}$ is monotonic and subhomogeneous.*

(b) The sum of two monotonic, subhomogeneous maps is monotonic and subhomogeneous.

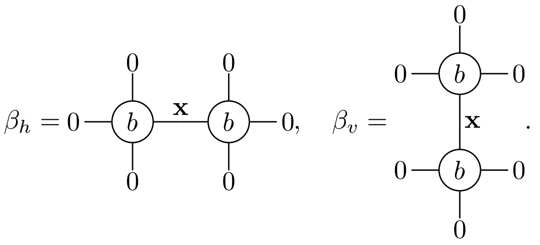
(c) The composition of two monotonic, subhomogeneous maps is monotonic and subhomogeneous.

(d) The product of a nonnegative monotonic, subhomogeneous map and a nonnegative monotonic real-valued function is monotonic and subhomogeneous.

We omit the proof. We now proceed to the construction of the 2x1 map on the nonlinear level, and of the associated master function.

2.6 Gauge transformation

We repeat the analysis of Section 2.3.2 at the full nonlinear level. We keep the definitions of g_h and g_v given in (2.6). In addition, we define the scalars



(2.35)

It is easily checked that with $\beta = \beta_h, \beta_v$

$$g^3 = -\beta g. \tag{2.36}$$

At the nonlinear level, Eqs. (2.4) for G_h, G_v are replaced by the exact equalities:

$$G = \mathbb{1}_V + g + \frac{1}{1+\sqrt{1-\beta}}\, g^2, \quad G^{-1} = \mathbb{1}_V - g + \frac{1}{1+\sqrt{1-\beta}}\, g^2. \tag{2.37}$$

The first equation is the definition of G. The second equation is easily checked using (2.36). We will be assuming that $|\beta| < 1$, and so G, G^{-1} are analytic functions of b.

Let us briefly discuss how this is related to [3]. There, the gauge transformation was done with $G_{\rm there} = \exp(g)$, while our definition corresponds to taking $G = \exp(cg)$ with $c = \arcsin\big(\sqrt{\beta}\big)/\sqrt{\beta} = 1 + O(\beta)$. The two expressions agree to linear order in g and give the same needed cancellations. We will use Eq. (2.37) as it is simpler for the quantitative analysis of higher-order terms.

We will apply G_h, G_v and their inverses to the horizontal, vertical legs of A, respectively, as in Fig. 1(a). We denote the resulting tensor by A_g. Partition function is preserved ([4], Sec.2.4.3):

$$\boxed{Z(A, \ell_x \times \ell_y) = Z(A_g, \ell_x \times \ell_y)}\,. \tag{2.38}$$

Note that A_g is not normalized; we will worry about normalization at the end. As discussed in Section 2.3.2, if we expand A_g in g, the first order in g part will cancel the sectors of b given in (2.7). We now discuss the main improvement with respect to [3]: a graphical language which allows to make this first-order cancellation explicit, and to seamlessly express the bounds on the remainders. The same language will be later used for disentangling, in a more complicated setting.

We start by writing each of G_h, G_v as sums: $G_h = \sum_{i=1}^5 h_i$, $G_v = \sum_{i=1}^5 v_i$. The tensors h_i, v_i, are given in Figs. 2 and 3; they represent various terms in (2.37). Note that we find it convenient to split the identity and g as a sum of two terms each. Furthermore we define:

$$h_i' = h_i,\ v_i = v_i' \text{ for } i = 1, 2, 5, \qquad h_i' = -h_i,\ v_i = -v_i' \text{ for } i = 3, 4. \tag{2.39}$$

Then $G_h^{-1} = \sum_{i=1}^5 h_i'$, and $G_v^{-1} = \sum_{i=1}^5 v_i'$.

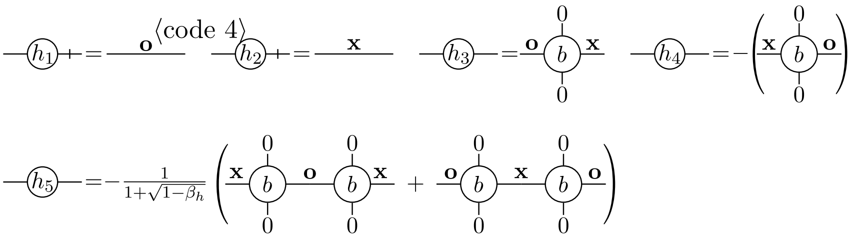


Figure 2: Definition of h_i tensors. The lines labeled **o** and **x** in the definitions of h_1 and h_2 represent orthogonal projectors onto the corresponding subspaces. The right legs of h_1, h_2 are distinguished, hence the tick mark.

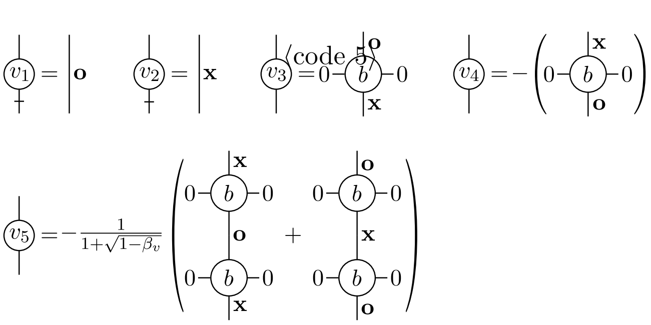
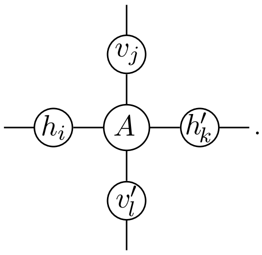


Figure 3: Definition of v_i tensors. The bottom leg of v_1, v_2 is distinguished.

Now A_g is given by the sum over i, j, k, l ranging from 1 to 5 of the following diagrams:

$$A_g = \sum_{ijkl} D^{(ijkl)}, \qquad D^{(ijkl)} := \tag{2.40}$$



⟨code 6⟩

As mentioned, there is some cancellation in A_g: terms first-order in g cancel a part of b. We make this explicit by defining a set of index combinations:

$$\mathcal{C} = \{2111, 4111, 1211, 1411, 1121, 1131, 1112, 1113\}. \tag{2.41}$$

The cancellation is now phrased by saying that the part of the sum in (2.40) over $ijkl \in \mathcal{C}$ is zero. For example the 2111 diagram picks up $b_{\mathbf{xooo}}$ while the 4111 diagram gives h_4 contracted with A_* (recall that $b_{0000} = 0$) which is exactly $-b_{\mathbf{xooo}}$. Similarly the other three pairs of diagrams cancel pairwise. We can therefore drop the canceled terms and restrict the sum (2.40) to $ijkl \notin \mathcal{C}$.

We now proceed to bounding A_g i.e. finding its hat-tensor $\widehat{A}_g$. It will be expressed in terms of a hat-tensor for A, which we write in the form

$$\widehat{A} = \widehat{A}_* + \widehat{b} \tag{2.42}$$

⟨code 8⟩

where $(\widehat{A}_*)_{\mathbf{oooo}} = 1$ and all other components are zero, while $(\widehat{b})_{\mathbf{oooo}} = 0$. We define hat or check-tensors for all other tensors in the diagrams $D^{(ijkl)}$:

- For $i = 3, 4, 5$ the tensors h_i, h'_i, v_i, v'_i are Hilbert-Schmidt. We define their hat-tensors from their expressions in Figs. 2,3. Using Lemma 2.1, these hat-tensors are expressed in terms of (contractions of) various components of $\widehat{b}$. Furthermore, the coefficients $1/(1+\sqrt{1-\beta})$ that appear in h_5, h'_5, v_5, v'_5 need to be replaced by appropriate upper bounds on their absolute values in terms of $\widehat{b}$. We have $|1/(1+\sqrt{1-\beta})| \leqslant 1/(1+\sqrt{1-|\beta|})$ for $|\beta| < 1$, and the r.h.s. is an increasing function of $|\beta|$. So the needed upper bounds are obtained by replacing β_h, β_v by upper bounds $\widehat{\beta}_h, \widehat{\beta}_v$ on the absolute values of β_h, β_v, given by[12]

$$\widehat{\beta}_h = \widehat{b}_{\mathbf{ooxo}}\, \widehat{b}_{\mathbf{xooo}}, \qquad \widehat{\beta}_v = \widehat{b}_{\mathbf{ooox}}\, \widehat{b}_{\mathbf{oxoo}}\,, \tag{2.43}$$

  as long as $\widehat{\beta}_h < 1$ and $\widehat{\beta}_v < 1$, which we assume to be the case.

- Tensors h_i, h'_i, v_i, v'_i, $i = 1, 2$ are orthogonal projectors, so we define their check-tensors (Remark 2.4). We declare the legs that are contracted with A to be distinguished, denoted with a tick mark in Figs. 2,3 for h_1, v_1, h_2, v_2.

  On the horizontal legs $V = \mathbf{o} \oplus \mathbf{d} \oplus \mathbf{u} \oplus \mathbf{r}$. The check-tensors $\check{h}_2, \check{h}'_2$ are 4×4 matrices with the **uu**, **dd**, **rr** elements equal to 1, and all other elements 0. The check-tensors $\check{h}_1, \check{h}'_1$ are 4×4 matrices with a single nonzero element **oo** equal to 1.

  On the vertical legs $V = \mathbf{o} \oplus \mathbf{x}$. The check-tensors $\check{v}_2$ $\check{v}'_2$ are 2×2 matrices with a single nonzero element **xx** equal to 1. The check-tensors $\check{v}_1$ $\check{v}'_1$ are 2×2 matrices with a single nonzero element **oo** equal to 1.

[12] Recall that $\mathbf{x} = \mathbf{d} \oplus \mathbf{u} \oplus \mathbf{r}$ on the horizontal legs; see Sec. 2.3.1. Contraction over such $\mathbf{x}$ implies a sum over $\mathbf{u}, \mathbf{d}, \mathbf{r}$. E.g. $\widehat{b}_{\mathbf{ooxo}}\, \widehat{b}_{\mathbf{xooo}}$ stands for $\sum_{a=\mathbf{u},\mathbf{d},\mathbf{r}} \widehat{b}_{\mathbf{oo}a\mathbf{o}}\, \widehat{b}_{a\mathbf{ooo}}$.

With the given definition of distinguished legs, all diagrams $D^{(ijkl)}$ are allowed contractions (Def. 2.3). So by Lemma 2.2 we obtain the hat-tensor for each diagram by replacing every tensor in the diagram by its hat or check-tensor, as defined above, and contracting these tensors. Summing the hat-tensors $\widehat{D}^{(ijkl)}$ over $ijkl \notin \mathcal{C}$, we obtain the hat-tensor $\widehat{A}_g$.

Here is a summary of our new graphical language. *We represent a full-tensor as a sum of diagrams. We render cancellations explicit by identifying and removing the set of diagrams which sum to zero. After that a hat-tensor is obtained by simply putting hats, or checks, on all the full-tensors involved. If there are coefficients in the diagrammatic expression for the full-tensor, we also need to replace these coefficients by bounds on their absolute values.*

There is one final step in our gauge transformation. The gauge transformations can change the 0000 component of our tensor by a small amount. To restore our normalization condition that the 0000 component is 1, we write $A_g = \mathcal{N}_1 A_1$, where $\mathcal{N}_1 = (A_g)_{0000}$ and A_1 is normalized, $A_1 = A_* + b_1$. Factoring out $\mathcal{N}_1$ from the partition function, we can rewrite (2.38) as:

$$\boxed{Z(A, \ell_x \times \ell_y) = \mathcal{N}_1^{\mathrm{Vol}} Z(A_1, \ell_x \times \ell_y)\,, \quad \mathrm{Vol} = \ell_x \ell_y}\,. \tag{2.44}$$

We now turn to finding a hat-tensor $\widehat{A}_1$, which we write as

$$\widehat{A}_1 = \widehat{A}_* + \widehat{b}_1 \tag{2.45}$$

where $(\widehat{A}_*)_{\mathbf{oooo}} = 1$ and has all other components are zero, while $(\widehat{b}_1)_{\mathbf{oooo}} = 0$. We also write $\widehat{A}_g$ in a similar form, separating the **oooo** component:

$$\widehat{A}_g = \widehat{\mathcal{N}}_1\, \widehat{A}_* + \widehat{b}_g, \qquad \widehat{\mathcal{N}}_1 := (\widehat{A}_g)_{\mathbf{oooo}}\,. \tag{2.46}$$

Since $A_1 = A_g/\mathcal{N}_1$, we can set $\widehat{b}_1 = \widehat{b}_g/|\mathcal{N}_1|_{plb}$ where $|\mathcal{N}_1|_{plb}$ is a *positive lower* bound for $|\mathcal{N}_1|$. Note that until now we only considered upper bounds on absolute values (or norms); e.g. $\widehat{\mathcal{N}}_1$ is an upper bound on $|\mathcal{N}_1|$. However a positive lower bound on $|\mathcal{N}_1|$ can be inferred without additional work, by the following argument. Our construction represents $\mathcal{N}_1 = 1 + n_1$ where n_1 is a sum of various contractions of b times various factors. The upper bound has the form $\widehat{\mathcal{N}}_1 = 1 + \widehat{n}_1$, where $\widehat{n}_1$ is an upper bound on $|n_1|$. Therefore, the needed positive lower bound has the form $|\mathcal{N}_1|_{plb} = 1 - \widehat{n}_1 = 2 - \widehat{\mathcal{N}}_1$. (It will be positive since we will impose $\widehat{\mathcal{N}}_1 < 2$.)

To summarize, the considerations of this section define the following functions:

$$\widehat{\beta}_h, \widehat{\beta}_v : \widehat{\mathbb{H}}_0 \to \mathbb{R}_{\geqslant 0}, \tag{2.47}$$

$$\widehat{\mathcal{N}}_1 : \{\widehat{b} \in \widehat{\mathbb{H}}_0 : \widehat{\beta}_h, \widehat{\beta}_v < 1\} \to \mathbb{R}_{\geqslant 0}, \tag{2.48}$$

$$\widehat{b}_1 : \widehat{\Omega}_1 \to \widehat{\mathbb{H}}_0, \quad \widehat{\Omega}_1 := \{\widehat{b} \in \widehat{\mathbb{H}}_0 : \widehat{\beta}_h < 1,\ \widehat{\beta}_v < 1,\ \widehat{\mathcal{N}}_1 < 2\}, \tag{2.49}$$

$$\mathcal{N}_1 : \Omega_1 \to \mathbb{C}, \qquad b_1 : \Omega_1 \to \mathbb{H}_0, \qquad \Omega_1 = \{b \in \mathbb{H}_0 : \exists\ \widehat{b} \text{ for } b \text{ such that } \widehat{b} \in \widehat{\Omega}_1\}. \tag{2.50}$$

Proposition 2.5. *(a) The sets on which $\widehat{\beta}_h$, $\widehat{\beta}_v$, $\widehat{\mathcal{N}}_1$, $\widehat{b}_1$ are defined are downward closed. $\widehat{\beta}_h$, $\widehat{\beta}_v$, $\widehat{b}$ are monotonic and subhomogeneous. $\widehat{\mathcal{N}}_1$ is monotonic.*

(b) The functions $\mathcal{N}_1$, b_1 are analytic on Ω_1.

(c) We have $b_1(0) = 0$, $\mathcal{N}_1(b) = 1 + O(b^2)$.

(d) If $\widehat{b}$ is a hat-tensor for b, then $\widehat{b}_1(\widehat{b})$ is a hat-tensor for $b_1(b)$. We have $\mathcal{N}_1 \neq 0$ on Ω_1, and the bound:

$$|\mathcal{N}_1(b) - 1| \leqslant \widehat{\mathcal{N}}_1(\widehat{b}) - 1\,. \tag{2.51}$$

Proof. (a) follows with the help of Lemma 2.4. Note that $1/(1 + \sqrt{1 - \widehat{\beta}_h})$, $1/(1 + \sqrt{1 - \widehat{\beta}_v})$ are nonnegative monotonic real functions of $\widehat{b}$ as long as $\widehat{\beta}_h, \widehat{\beta}_v < 1$, so that Lemma 2.4(d) can be applied. (b) follows from the fact that on Ω_1 we have $|\beta_h|, |\beta_v| < 1$, $|\mathcal{N}_1| > 0$, so that $1/(1+\sqrt{1-\beta_h})$, $1/(1+\sqrt{1-\beta_v})$, $1/\mathcal{N}_1$ are analytic. Multiplying by these factors were the only non-polynomial operations in the definitions of $\mathcal{N}_1$, b_1. (c) follows from definitions. (d) is just a restatement of things already discussed. □

2.7 Disentangling and splitting

Here we will repeat the linearized discussion in Section 2.3.3 at the fully nonlinear level. Recall that disentangling and splitting is applied to the tensor $A_1 = A_* + b_1$ which is the end result of the gauge transformation step. As in Section 2.3.3, we rename A_1 as A and b_1 as b. This will avoid writing index 1 on many tensors below. We will revert to A_1 and b_1 in Section 2.7.6.

Disentangling and splitting is represented in Eq. (2.8), which is a copy of Fig. 1(b) up to the described renaming. So, we divide the lattice into horizontal 2×1 blocks and contract the two A's in each block. Our disentangler D will be a tensor with two legs on top and two on the bottom. We think of it as a matrix acting from the space corresponding to the two bottom legs to the space corresponding to the top two legs. In our network we insert $I = D^{-1}D$ on each pair of vertical legs between two blocks. We think of the D above a block and the D^{-1} below the same block as being in that block. The resulting block is shown in the left side of Eq. (2.8). (The right side of the figure is for later use.)

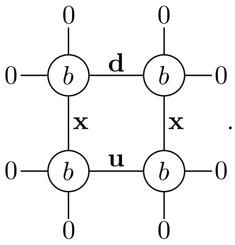
We want to choose the disentangler D to disentangle the contraction of two A's in the block. The major part of this contraction comes from the 0 index, but $\mathbf{x}$ indices also contribute. Crudely speaking, disentangling means reducing the contribution of $\mathbf{x}$ indices to the contraction. Concretely, we will construct D to cancel the diagrams shown in Eq. (2.12). In this figure the contracted legs are never in $\mathbf{o}$, so A_* does not contribute and so we have replaced A by b. The reason for cancelling these particular diagrams was explained in the linearized analysis. A discussion of the type of disentangler we are using can be found in Section 2.4.4 of [4]. In particular Remark 2.6 discusses the origin of the name disentangler.[13]

If we group the top two legs together and the bottom two legs together in Eq. (2.12), then the problem of cancelling these diagrams is analogous to how we cancelled certain diagrams in the original A with a gauge transformation. So our disentangler D is quite similar to the gauge transformation G defined in Eq. (2.37). It is defined by (cf. (2.10)):

$$D = \mathbb{1}_{V\otimes V} + X + \frac{1}{1+\sqrt{1-\alpha}}X^2, \qquad D^{-1} = \mathbb{1}_{V\otimes V} - X + \frac{1}{1+\sqrt{1-\alpha}}X^2, \tag{2.52}$$

where X is given in Eq. (2.11), and

$\alpha =$



. (2.53)

Below we will make sure that $|\alpha| < 1$, so that (2.52) are well defined and analytic in b. Here, X should be considered as a matrix acting from the space of its bottom legs to the space of its top legs. Similarly to the gauge transformation case, the formula for D^{-1} works because $X^3 = -\alpha X$. If we compute the contraction of D, D^{-1} and the two A's in Eq. (2.8) to first order in X, we see that the diagrams from Eq. (2.12) are cancelled.

[13] Ref. [9] introduced disentanglers into numerical algorithms of tensor network RG. Ref. [3] first used disentanglers in a rigorous tensor RG study. For a closely related use of disentanglers in quantum lattice models see [10].

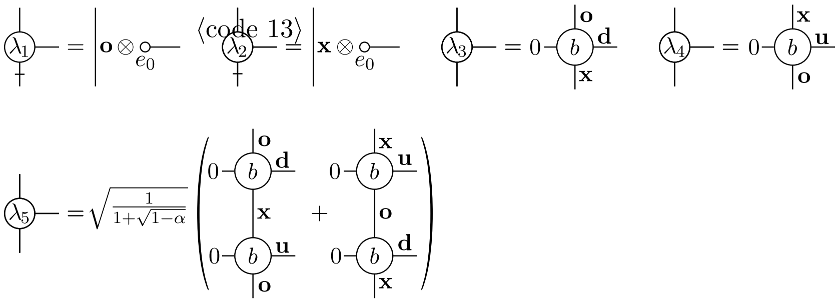


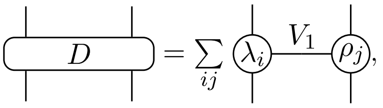
Figure 4: Definition of λ_i tensors. Here, λ_1 and λ_2 are Kronecker products of the shown projectors with e_0, the basis vector of $\mathbf{o}$, a 1-tensor denoted by the circle with label e_0.

2.7.1 Decomposition of disentangler

The next step is to decompose the disentangler D into a sum of terms. This will be extremely helpful to accomplish two goals. One goal is to represent the contraction of two A's with D and D^{-1} as the contraction of two new tensors L and R as shown in Eq. (2.8). We refer to this step as "splitting." The second goal is to explicitly realize the discussed cancellation (see Sec. 2.7.4 below). The l.h.s. of Eq. (2.8) will become a sum of terms, some of which will cancel each other exactly.

So we introduce $(\lambda_i)_{i=1}^5$ tensors in Fig. 4. The right leg of these tensors lives in the Hilbert space $V_1 := V \oplus (V \otimes V)$. The sectors of V_1 have the form a or $a \otimes b$ where a, b are sectors of V. All tensor elements not shown in Fig. 4 are assumed zero. In particular, tensors λ_i with $i = 1, 2, 3, 4$ vanish when the right leg is in the $V \otimes V$ subspace of V_1, while λ_5 vanishes when the right leg is in the V subspace of V_1.

We also define ρ_i tensors in an analogous way but with all sector labels reflected with respect to a vertical line and tensors ρ_4, ρ_5 **acquiring minus signs** (this is needed to reproduce minus signs of some diagrams in X and X^2). With these definitions it's easy to check that



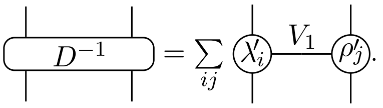
(2.54)

where we put V_1 label on the contracted leg to emphasize that it lives in a different Hilbert space.

Some comments are in order:

- Since λ_5 and ρ_5 are the only ones which have horizontal legs in the $V \otimes V$ part of V_1, they contract to each other but not with any other tensor. Their contraction reproduces the X^2 term in (2.52) including the prefactor.
- Since λ_i, ρ_j with $i, j = 1, 2$ are the only ones which have e_0 horizontal legs, they only contract among themselves. Their contraction reproduces exactly the $\mathbb{1}_{V\otimes V}$ term in D.
- Finally, λ_3 contracts only with ρ_3, and λ_4 only with ρ_4. These contractions reproduce the X term in (2.52).

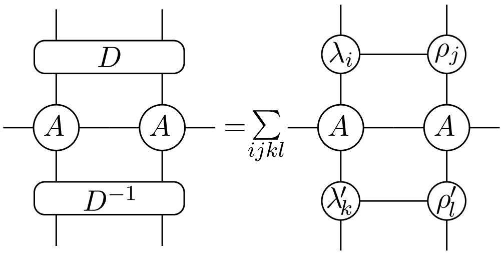
We also define λ'_i and ρ'_i in an analogous way so that:



(2.55)

They only differ from λ_i and ρ_i by some signs to account for the opposite sign of X in D^{-1}; we skip the exact definition.

Substituting these expressions for D and D^{-1} into the left side of Eq. (2.8) we get:



(2.56)

2.7.2 Channels

We now proceed to the splitting, i.e. to defining the tensors L, R in the r.h.s. of (2.8). Crudely speaking, splitting will be performed by cutting each diagram in (2.56) across the three horizontal bonds and putting the left side into L and the right side into R. We represented the disentangler D as a sum of contractions of λ_i and ρ_j so that we could perform this cutting. Further tricks will accompany the cutting, whose role will be explained in due course.

As the first step we write the r.h.s. of Eq. (2.56) as $\sum_q H_q$ with the five terms H_q, $q \in \{z, \mathbf{xx}, \mathbf{ox}, \mathbf{xo}, \mathbf{oo}\}$ defined in Fig. 5. We will refer to these five tensors as channels. For the $\mathbf{xx}, \mathbf{ox}, \mathbf{xo}, \mathbf{oo}$, the name refers to the sectors on the horizontal legs. The z ("zero") channel is different in that it is the horizontal contracted bonds which are restricted to $\mathbf{o}$. These channels are the nonlinear counterparts of the five channels with the same names considered in the linearized analysis in Section 2.3.6.

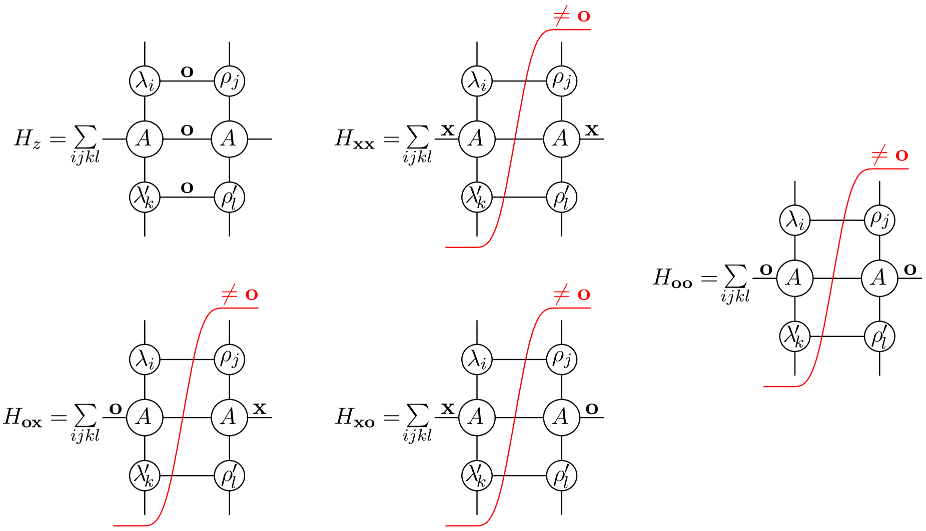
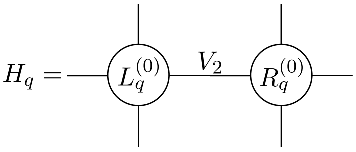

$H_z = \sum_{ijkl}$
o
o
o
$H_{\mathbf{xx}} = \sum_{ijkl}$
x
x
$\neq$ o
$H_{\mathbf{oo}} = \sum_{ijkl}$
o
o
$\neq$ o
$H_{\mathbf{ox}} = \sum_{ijkl}$
o
x
$\neq$ o
$H_{\mathbf{xo}} = \sum_{ijkl}$
x
o
$\neq$ o
λ_i
ρ_j
A
λ'_k
ρ'_l

Figure 5: Definition of channels. A line cutting through the bonds or legs with "$\neq$ **o**" label means that we forbid all these bonds or legs to be simultaneously in the **o** sector. In this particular case, this is the same as inserting there a projector to the orthogonal complement of $\mathbf{o} \otimes \mathbf{o} \otimes \mathbf{o}$ in $V_1 \otimes V \otimes V_1$.

Next, we write each channel H_q in the natural way as a contraction of two tensors denoted $L_q^{(0)}$

and $R_q^{(0)}$:[14]



(2.57)

The involved steps are as follows (see Fig. 6 for the explicit definitions):

⟨code 15⟩

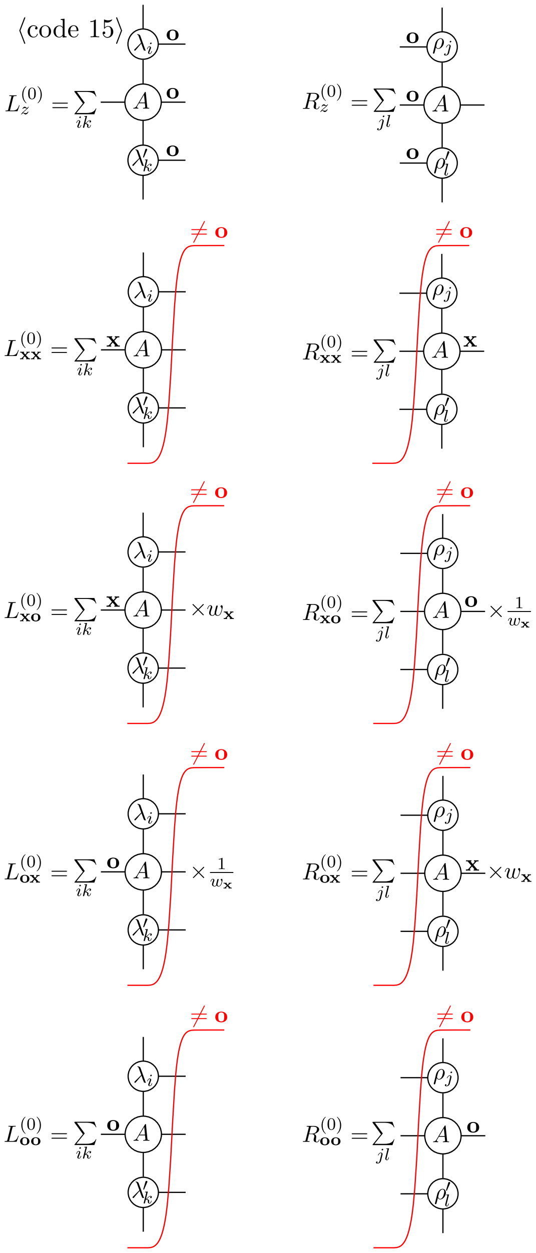

$L_z^{(0)} = \sum_{ik}$
λ_i
A
$\lambda_k^\prime$
$\mathbf{o}$
$R_z^{(0)} = \sum_{jl}$
ρ_j
$\rho_l^\prime$
$L_{\mathbf{xx}}^{(0)} = \sum_{ik}$
$\mathbf{x}$
$\neq \mathbf{o}$
$R_{\mathbf{xx}}^{(0)} = \sum_{jl}$
$L_{\mathbf{xo}}^{(0)} = \sum_{ik}$
$\times w_{\mathbf{x}}$
$R_{\mathbf{xo}}^{(0)} = \sum_{jl}$
$\times \frac{1}{w_{\mathbf{x}}}$
$L_{\mathbf{ox}}^{(0)} = \sum_{ik}$
$R_{\mathbf{ox}}^{(0)} = \sum_{jl}$
$L_{\mathbf{oo}}^{(0)} = \sum_{ik}$
$R_{\mathbf{oo}}^{(0)} = \sum_{jl}$

Figure 6: The definition of $L_q^{(0)}$, $R_q^{(0)}$ tensors.

[14] Extra manipulations will be needed before $L_q^{(0)}$ and $R_q^{(0)}$ are assembled into L and R, hence the superscript (0).

- We cut the three horizontal bonds in H_q. The cut bonds become horizontal legs of $L_q^{(0)}$ and $R_q^{(0)}$, living in $V_2 := V_1 \otimes V \otimes V_1$. (We assume the ordering convention for the triple of bonds top, middle, bottom.) We emphasized this by V_2 label in Eq. (2.57).

- We factor $\sum_{i,j,k,l=1}^{5} = \sum_{i,k=1}^{5} \sum_{j,l=1}^{5}$ in the definition of H_q, with ik the indices of λ, λ' and jl the indices of ρ, ρ'. The sum over ik goes into $L_q^{(0)}$, the sum over jl into $R_q^{(0)}$.

- Furthermore, for the **ox** and **xo** channels we redistribute the weight of the diagram by the introduction of a parameter $w_{\mathbf{x}} > 0$ which multiples one of the factors $L_q^{(0)}$, $R_q^{(0)}$ and divides the other. We saw in Section 2.3.6 why this parameter is useful.

2.7.3 Auxiliary space

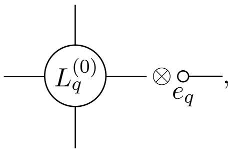
So we represented H_q as a contraction of $L_q^{(0)}$ and $R_q^{(0)}$ for each q. We now wish to write the r.h.s. of Eq. (2.56), which equals $\sum_q H_q$, as a contraction of $\sum_q L_q$ with $\sum_{q'} R_{q'}$. We could try to take $L_q \overset{?}{=} L_q^{(0)}$, $R_q \overset{?}{=} R_q^{(0)}$, but this would not work. The diagonal terms $q = q'$ produce the five channels (good), but the cross terms $q \neq q'$ produce terms which do not exist in $\sum_q H_q$ (bad).

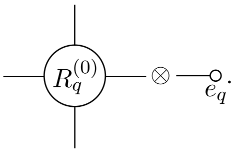
To get rid of the cross terms, we introduce an auxiliary seven-dimensional Hilbert space V_{aux} spanned by the following orthonormal basis:

$$V_{\text{aux}} = \text{span}(e_z, e_{\mathbf{xx}}, e_{\mathbf{ox}}, e_{\mathbf{xo}}, e_{\mathbf{oo}}, e'_{\mathbf{oo}}, e''_{\mathbf{oo}}). \tag{2.58}$$

The vectors $e_z, e_{\mathbf{xx}}, e_{\mathbf{ox}}, e_{\mathbf{xo}}$ will be used for the $z, \mathbf{xx}, \mathbf{ox}, \mathbf{xo}$ channels, respectively, while the last three vectors will be used for the **oo** channel.

Now, for $q = z, \mathbf{xx}, \mathbf{ox}, \mathbf{xo}$ (the **oo** channel is considered separately below), we define

$L_q = L_q^{(0)} \otimes e_q =$ 

, $R_q = R_q^{(0)} \otimes e_q =$ 

. (2.59)

Tensoring with a new vector e_q is a formal realization of the idea of introducing sector copies, used in the linearized analysis in Section 2.3.6. We will stick to the language of tensoring from now on.

We think of L_q (resp. R_q) as a 4-tensor, whose left (resp. right) leg lives in the Hilbert space

$$V_3 := V_2 \otimes V_{\text{aux}} = V_1 \otimes V \otimes V_1 \otimes V_{\text{aux}}. \tag{2.60}$$

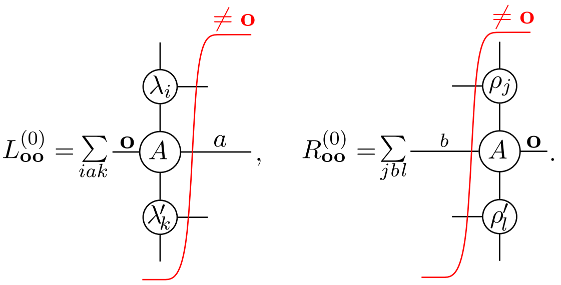
Tensors L_q and R_q so defined have the same contraction as $L_q^{(0)}$ and $R_q^{(0)}$, i.e. H_q. Furthermore, contractions between L_q and $R_{q'}$, $q' \neq q$, vanish. Therefore, there are no unwanted cross terms, and we have ($\overset{V_3}{\text{---}}$ denotes contraction over V_3):

$$H_z + H_{\mathbf{xx}} + H_{\mathbf{ox}} + H_{\mathbf{xo}} = (L_z + L_{\mathbf{xx}} + L_{\mathbf{ox}} + L_{\mathbf{xo}}) \overset{V_3}{\text{---}} (R_z + R_{\mathbf{xx}} + R_{\mathbf{ox}} + R_{\mathbf{xo}}). \tag{2.61}$$

2.7.4 The oo channel

Just as in the linearized analysis, this needs a special treatment as this is the channel where we have exact cancellations resulting from our choice of the disentangler D, see (2.12) and (2.11).

As a first step, we rewrite $L^{(0)}_{\mathbf{oo}}$ and $R^{(0)}_{\mathbf{oo}}$ in Fig. 6, by splitting the middle horizontal legs into sectors a, summed over $\mathbf{o}, \mathbf{d}, \mathbf{u}, \mathbf{r}$:



⟨code 16⟩ (2.62)

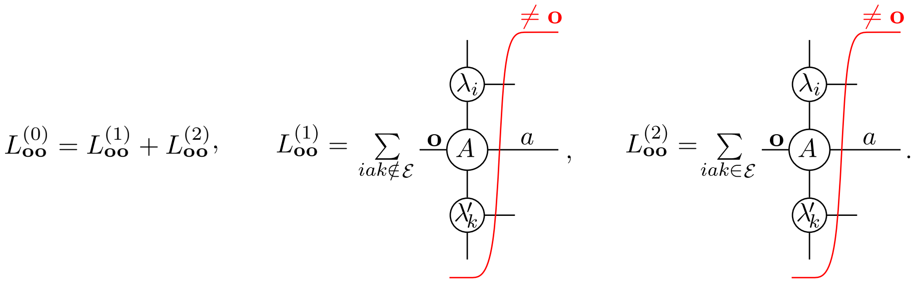
Now define the following set of groups of indices:

$$\mathcal{E} = \{(2, \mathbf{u}, 1), (1, \mathbf{d}, 2), (4, \mathbf{o}, 1), (1, \mathbf{o}, 3)\}\,. \tag{2.63}$$

⟨code 17⟩

Here is a crucial observation: In the contraction of $L_{\mathbf{oo}}$ with $R_{\mathbf{oo}}$ defined by (2.62), if we restrict the sums to those terms with $(i, a, k) \in \mathcal{E}$ and $(j, b, l) \in \mathcal{E}$, then this partial sum is zero. (The first two elements of $\mathcal{E}$ reproduce the two diagrams in (2.12) and the second two elements of $\mathcal{E}$ reproduce the same diagrams with a minus sign.) This is the explicit realization of the cancellation mentioned at the beginning of Sec. 2.7.1.

In light of this observation let us split $L^{(0)}_{\mathbf{oo}} = L^{(1)}_{\mathbf{oo}} + L^{(2)}_{\mathbf{oo}}$ where the two parts correspond to restricting the summation in (2.62) to $(i, k, a) \notin \mathcal{E}$ (resp. $(i, k, a) \in \mathcal{E}$):



$L^{(0)}_{\mathbf{oo}} = L^{(1)}_{\mathbf{oo}} + L^{(2)}_{\mathbf{oo}}$,
$L^{(1)}_{\mathbf{oo}} = \sum_{iak\notin\mathcal{E}}$
$\neq \mathbf{o}$
λ_i
$\mathbf{o}$
A
a
λ'_k
,
$L^{(2)}_{\mathbf{oo}} = \sum_{iak\in\mathcal{E}}$
$\neq \mathbf{o}$
λ_i
$\mathbf{o}$
A
a
λ'_k
.

(2.64)

We also split $R^{(0)}_{\mathbf{oo}} = R^{(1)}_{\mathbf{oo}} + R^{(2)}_{\mathbf{oo}}$ in the same fashion.

Recall that the contraction of $L^{(0)}_{\mathbf{oo}}$ with $R^{(0)}_{\mathbf{oo}}$ produces the zero channel. Note that the contraction of $L^{(2)}_{\mathbf{oo}}$ with $R^{(2)}_{\mathbf{oo}}$ is zero since this contraction is precisely the sum over $\mathcal{E}$ which contains the terms that cancel. Dropping this vanishing contraction, we are left with the decomposition of $\mathbf{oo}$ into the sum of three contractions over V_2:

$$H_{\mathbf{oo}} = L^{(1)}_{\mathbf{oo}} \overset{V_2}{\text{---}} R^{(1)}_{\mathbf{oo}} \;+\; L^{(2)}_{\mathbf{oo}} \overset{V_2}{\text{---}} R^{(1)}_{\mathbf{oo}} \;+\; L^{(1)}_{\mathbf{oo}} \overset{V_2}{\text{---}} R^{(2)}_{\mathbf{oo}}\,. \tag{2.65}$$

Given this equation, we now define $L_{\mathbf{oo}}$ and $R_{\mathbf{oo}}$ so that

$$H_{\mathbf{oo}} = L_{\mathbf{oo}} \overset{V_3}{\text{---}} R_{\mathbf{oo}}. \tag{2.66}$$

To achieve this we use the auxiliary vector trick from Sec. 2.7.3. We define (For the first reading replace $w_{\mathbf{o}}$ and P_Y by 1; see immediately below for the explanation of these factors.)

$$\begin{aligned} L_{\mathbf{oo}} &= L^{(1)}_{\mathbf{oo}} \otimes \underset{e_{\mathbf{oo}}}{\circ\!\!\text{---}} + L^{(1)}_{\mathbf{oo}} P_Y \otimes \underset{e'_{\mathbf{oo}}}{\circ\!\!\text{---}} \times w_{\mathbf{o}} + L^{(2)}_{\mathbf{oo}} \otimes \underset{e''_{\mathbf{oo}}}{\circ\!\!\text{---}} \times \tfrac{1}{w_{\mathbf{o}}}\,, \\ R_{\mathbf{oo}} &= R^{(1)}_{\mathbf{oo}} \otimes \underset{e_{\mathbf{oo}}}{\text{---}\!\!\circ} + P_Y R^{(1)}_{\mathbf{oo}} \otimes \underset{e''_{\mathbf{oo}}}{\text{---}\!\!\circ} \times w_{\mathbf{o}} + R^{(2)}_{\mathbf{oo}} \otimes \underset{e'_{\mathbf{oo}}}{\text{---}\!\!\circ} \times \tfrac{1}{w_{\mathbf{o}}}\,. \end{aligned} \tag{2.67}$$

In words, we consider the sums of Kronecker products of the left (resp. right) factors in (2.65) with the auxiliary vectors $e_{\mathbf{oo}}$, $e'_{\mathbf{oo}}$, $e''_{\mathbf{oo}}$, to eliminate the cross terms which would appear otherwise.[15]

We now explain $w_{\mathbf{o}}$ and P_Y in (2.67). The $w_{\mathbf{o}} > 0$ is a reweighting parameter already discussed in Section 2.3.6. As for P_Y, consider the following three subspaces in V_2:

$$\mathbf{o} \otimes (\mathbf{u} \oplus \mathbf{d}) \otimes \mathbf{o}, \quad \mathbf{u} \otimes \mathbf{o} \otimes \mathbf{o}, \quad \mathbf{o} \otimes \mathbf{o} \otimes \mathbf{d}\,. \tag{2.68}$$

Let Y be their direct sum and P_Y the corresponding orthogonal projector. It's easy to check that the horizontal legs of $L^{(2)}_{\mathbf{oo}}$ and $R^{(2)}_{\mathbf{oo}}$ live in Y. Hence the result does not change if we multiply the tensors contracted with $L^{(2)}_{\mathbf{oo}}$ and $R^{(2)}_{\mathbf{oo}}$ by P_Y, which is what we did. As a matter of fact, this improvement is not essential but it leads to slightly better numerical estimates, so we implement it.

2.7.5 Definition of L and R

So far we represented the **oo** channel as a contraction of $L_{\mathbf{o}}$ and $R_{\mathbf{o}}$, Eq. (2.66), and the sum of 4 other channels as a contraction of the sum of L_q and R_q over those channels, Eq. (2.61). Given these and footnote 15, we have

$$\sum_q H_q = L \overset{V_3}{\text{---}} R, \qquad L := \sum_q L_q, \qquad R := \sum_q R_q, \tag{2.69}$$

where the sum is over all 5 channels.

Recall that $\sum_q H_q$ equals the r.h.s. of (2.56). So Eq. (2.69) achieves the splitting. We will do one more minor step, whose purpose is to "erase" some unnecessary structure present in V_3.

Consider any isometric map $\varkappa$ from V_3 onto $V = \mathbf{o} \oplus \mathbf{x}$, which has the following property:[16]

$$\varkappa : e_0 \otimes e_0 \otimes e_0 \otimes e_z \mapsto e_0 \tag{2.70}$$

(and hence $\varkappa$ maps the orthogonal complement $(e_0 \otimes e_0 \otimes e_0 \otimes e_z)^\perp$ onto $\mathbf{x}$). We define tensors $L^\varkappa$ and $R^\varkappa$ contracting the V_3 leg of L and R with $\varkappa$, i.e.

$$L^\varkappa = L \overset{V_3}{\text{---}} \varkappa, \quad R^\varkappa = \varkappa \overset{V_3}{\text{---}} R\,. \tag{2.71}$$ ⟨code 20⟩

Since $\varkappa$ is an isometry, the contraction of $L^\varkappa$ and $R^\varkappa$ over V is the same as of L and R over V_3:

$$\sum_q H_q = L^\varkappa \overset{V}{\text{---}} R^\varkappa\,, \tag{2.72}$$

This is the final form of splitting, which will be used below.

Let us split the tensors $L^\varkappa$ and $R^\varkappa$ into A_* plus remainders. The component $L^\varkappa_{0000}$ only has a contribution from L_z, Eq. (2.75). Using the definitions of λ_i and λ'_i in Fig. 4 and the definition of $L^{(0)}_z$ in Fig. 6, we see that only the term with $i = 1, k = 1$ in the latter figure contributes. Then, using $A = A_* + b$, $b_{0000} = 0$, we get $L^\varkappa_{0000} = 1$. Similarly, $R^\varkappa_{0000} = 1$. Thus we can write

$$L^\varkappa = A_* + b_L, \qquad R^\varkappa = A_* + b_R, \qquad (b_L)_{0000} = (b_R)_{0000} = 0. \tag{2.73}$$

Remark 2.5. Questions and answers.
Q: The presented strategy for solving the equation in Eq. (2.8) looks complicated. In numerical algorithms of tensor RG, one solves similar equations using singular value decomposition (SVD).

[15] There are also no cross terms between $L_{\mathbf{oo}}$, $R_{\mathbf{oo}}$ and the L_q, R_q tensors defined in (2.59) for the other 4 channels.

[16] Since V_3 and V are separable Hilbert spaces, such a map $\varkappa$ exists and is vastly non-unique. We will not need its explicit expression. It can be obtained e.g. as an identity matrix with respect to a numbering of the orthonormal basis elements of the two Hilbert spaces.

Why can't we simply define L and R by performing an SVD of the r.h.s. of Eq. (2.8) as USV, with U, V unitary and S nonnegative diagonal, and then defining $L_{\rm svd} = US^{1/2}$ and $R_{\rm svd} = S^{1/2}V$?
A: This would not work for us for several reasons. First, tensors $L_{\rm svd}$ and $R_{\rm svd}$ are not guaranteed to be Hilbert-Schmidt in our infinite-dimensional setting. Second, since SVD involves matrix diagonalization, it would be nontrivial to estimate tensor elements of $L_{\rm svd}$ and $R_{\rm svd}$ in terms of tensor elements of A. In contrast, our definition of L and R involves nothing more than reshuffling and contracting tensor pieces, and the needed estimates will follow easily.
Q: What is the intuition behind the need to introduce the "big" Hilbert space V_3?
A: The equation in Eq. (2.8) has to be satisfied for arbitrary values of indices on 6 uncontracted legs. It is not so surprising that to solve this equation, the space of contracted indices should be in some sense "bigger" than V. In our infinite-dimensional setting, "bigger" means a complicated tensor product. In finite dimensions "bigger" would be literal, as the dimension of V_3 would have to be at least χ^3 where χ is the dimension of V.
Q: Why erase almost all structure of V_3? What is special about the vector $e_0 \otimes e_0 \otimes e_0 \otimes e_z$?
A: We are studying the RG stability of the high-T fixed point. All tensors are split as A_* plus remainders, and we only need to control the norm of the remainders. The basis vector $e_0 \otimes e_0 \otimes e_0 \otimes e_z$ is the only component which contributes to the A_* part of $L^\varkappa$ and $R^\varkappa$. It's not important how the other components are distributed over the tensor factors of V_3, so they can be lumped into $\mathbf{x}$.
Q: Why are we keeping only two sectors $\mathbf{o}$, $\mathbf{x}$ on the contracted horizontal leg of $L^\varkappa$, $R^\varkappa$, and not four as for the horizontal legs of A?
A: The subsequent 90 degree rotation will turn this horizontal leg into a vertical one.

2.7.6 Hat-tensor estimates

We will now derive the hat-tensors $\widehat{L}^\varkappa, \widehat{R}^\varkappa$ for $L^\varkappa, R^\varkappa$. As with the gauge transformation, our graphical language makes this almost automatic.

- We define the hat-tensors for $\lambda_3, \lambda_4, \lambda_5$ by putting hats on the b's in Fig. 4. For λ_5, we also need to bound $1/(1+\sqrt{1-\alpha})^{1/2}$. We define an upper bound for $|\alpha|$:

$$\widehat{\alpha} = \widehat{b}_{\mathbf{oodx}}\, \widehat{b}_{\mathbf{doox}}\, \widehat{b}_{\mathbf{uxoo}}\, \widehat{b}_{\mathbf{oxuo}}, \qquad (2.74) \qquad \langle\text{code } 21\rangle$$

  We will assume that $\widehat{\alpha} < 1$. We then have $|1/(1+\sqrt{1-\alpha})^{1/2}| \leqslant 1/(1+\sqrt{1-|\alpha|})^{1/2} \leqslant 1/(1+\sqrt{1-\widehat{\alpha}})^{1/2}$. We use the latter bound in the definition of the hat-tensor.

- To obtain $\widehat{\lambda}'_i, \widehat{\rho}_i, \widehat{\rho}'_i$, $i = 3, 4, 5$, we first drop all the sign factors appearing in the definitions of the corresponding full-tensors and then repeat the same procedure as for $\lambda_3, \lambda_4, \lambda_5$.

- We define the check-tensors for λ_1, λ_2 by replacing the projector to $\mathbf{o}$ with the 2×2 matrix with a single nonzero element $\mathbf{oo}$ equal to 1, the projector to $\mathbf{x}$ with the 2×2 matrix with a single nonzero element $\mathbf{xx}$ equal to 1, and e_0 with a 4-dimensional vector with a single nonzero element $\mathbf{o}$ equal to 1.

- To obtain $\check{\lambda}'_i, \check{\rho}_i, \check{\rho}'_i$, $i = 1, 2$, we first drop all the sign factors appearing in the definitions of the corresponding full-tensors and then perform the same replacements as for λ_1, λ_2.

- Hat-tensors for the orthonormal basis vectors of $V_{\rm aux}$ can be taken equal to the vectors themselves. (I.e. we split $V_{\rm aux}$ into 7 one-dimensional sectors, one for each of its basis vectors.)

- It's now easy to check that the diagrams defining tensors L_q and R_q are allowed, and to define hat-tensors $\widehat{L}_q$ and $\widehat{R}_q$ using Lemma 2.2. Summing those over q we get hat-tensors $\widehat{L}$ and $\widehat{R}$.

Note that V_3 has $N_3 = N_1 \times 4 \times N_1 \times 7$ sectors, where $N_1 = 4 + 4 \times 4 = 20$ is the number of sectors of V_1. So $\widehat{L}$ and $\widehat{R}$ are 4-tensors with $4 \times 2 \times N_3 \times 2$ components.[17] On the other hand, the hat-tensors $\widehat{L}^{\varkappa}$ and $\widehat{R}^{\varkappa}$ will only have $4 \times 2 \times 2 \times 2$ components.

There are two strategies to get $\widehat{L}^{\varkappa}$ and $\widehat{R}^{\varkappa}$ from $\widehat{L}$ and $\widehat{R}$. The simplest one would be to define a check-tensor $\check{\varkappa}$ and apply Lemma 2.2. Below we describe a different strategy, which leads to much better hat-tensors. It takes advantage of the important fact that $\varkappa$, being an isometry, maps sectors of V_3 to mutually orthogonal subspaces of V.

Defining hat-tensor for $L^{\varkappa}$ means bounding the HS norm of its restriction $(L^{\varkappa})_{abcd}$. We consider the two cases $c = \mathbf{o}, \mathbf{x}$ separately. Denote by $\mathbf{s}_0$ the sector of V_3 spanned by $e_0 \otimes e_0 \otimes e_0 \otimes e_z$. We have, for any a, b, d

$$L^{\varkappa}_{ab\mathbf{o}d} = L_{ab\mathbf{s}_0 d} = (L_z)_{ab\mathbf{s}_0 d}, \tag{2.75}$$

where the first equation is by definition of $L^{\varkappa}$ and the second follows because L_z is the only tensor among L_q's whose right leg lives in $\mathbf{s}_0$. Therefore we obtain the hat-tensor:

$$\widehat{L}^{\varkappa}_{ab\mathbf{o}d} = (\widehat{L}_z)_{ab\mathbf{s}_0 d}, \tag{2.76}$$

⟨code 22⟩

Let now $(\mathbf{s}_r)_{r=1}^{N_3-1}$ be the remaining used sectors of V_3 in an arbitrary order. We have, for any fixed indices $i \in a, j \in b, l \in d$,

$$L^{\varkappa}_{ij\mathbf{x}l} = \sum_{r=1}^{N_3-1} L_{ij\mathbf{s}_r l}\relbar\!\!\relbar\varkappa_{\mathbf{s}_r \mathbf{x}}\,. \tag{2.77}$$

The terms in the r.h.s. are vectors in the $\mathbf{x}$ subspace of V which are mutually orthogonal for different values of r. This follows from the facts that sectors of V_3 are mutually orthogonal and that $\varkappa$ being an isometry maps orthogonal vectors to orthogonal vectors. So the square of the norm of the l.h.s. equals the sum over r of squares of the norms of the terms in the r.h.s. Summing the resulting equation over i, j, l and taking the square root, we obtain:

$$\widehat{L}^{\varkappa}_{ab\mathbf{x}d} = \left(\sum_{r=1}^{N_3-1} (\widehat{L}_{ab\mathbf{s}_r d})^2 \right)^{1/2}. \tag{2.78}$$

⟨code 23⟩

Note that for any given $\mathbf{s}_r$, the term in the r.h.s. of this equation receives contributions from one of the three hat-tensors $\widehat{L}_{\mathbf{xx}}$, $\widehat{L}_{\mathbf{ox}}$, $\widehat{L}_{\mathbf{xo}}$, or one of the three pieces of $\widehat{L}_{\mathbf{oo}}$ in the r.h.s. of (2.67). The $\widehat{L}_z$ does not contribute here.

Remark 2.6. Because of several terms combined in quadrature as in (2.78), $\widehat{L}^{\varkappa}$ is not an analytic function of $\widehat{b}$. On the other hand $L^{\varkappa}$ is analytic function of b. There is no contradiction of course.

The discussion for $\widehat{R}^{\varkappa}$ is completely analogous; we omit the details.

Recall that $L^{\varkappa}$ and $R^{\varkappa}$ are normalized and were split into A_* plus remainders as in Eq. (2.73). So their hat-tensors obtained as above can be written as

$$\widehat{L}^{\varkappa} = \widehat{A}_* + \widehat{b}_L, \qquad \widehat{R} = \widehat{A}_* + \widehat{b}_R, \qquad (\widehat{b}_L)_{\mathbf{oooo}} = (\widehat{b}_R)_{\mathbf{oooo}} = 0. \tag{2.79}$$

To summarize, in this section we defined the following functions:[18]

$$\widehat{\alpha} : \widehat{\mathbb{H}}_0 \to \mathbb{R}_{\geqslant 0}, \tag{2.80}$$

$$\widehat{b}_L, \widehat{b}_R : \widehat{\Omega}_{LR} \to \widehat{\mathbb{H}}_0, \quad \widehat{\Omega}_{LR} = \{\widehat{b} \in \widehat{\mathbb{H}}_0 : \widehat{\alpha} < 1\}, \tag{2.81}$$

$$b_L, b_R : \Omega_{LR} \to \mathbb{H}_0, \quad \Omega_{LR} = \{b \in \mathbb{H}_0 : \exists\, \widehat{b} \text{ for } b \text{ such that } \widehat{b} \in \widehat{\Omega}_{LR}\}\,. \tag{2.82}$$

[17] Many of these components are zero. Only 5 sectors of V_1 are actually used by λ and ρ ($\mathbf{o}$, $\mathbf{d}$, $\mathbf{u}$, $\mathbf{d} \otimes \mathbf{u}$, $\mathbf{u} \otimes \mathbf{d}$). So the number of used V_3 sectors is only $5 \times 4 \times 5 \times 7$. In the computer code that we use to compute the hat-tensors this is taken into account to speed up the computations.

[18] As noted above, full-tensors b_L, b_R have only two sectors $\mathbf{o}, \mathbf{x}$ on their contracted legs, not 4 sectors like b. Not to introduce further notation, we will continue to use $\mathbb{H}_0$ to denote the space in which these full-tensors live, as well as $\widehat{\mathbb{H}}_0$ for the space of their hat-tensors. Hopefully this will not lead to a confusion.

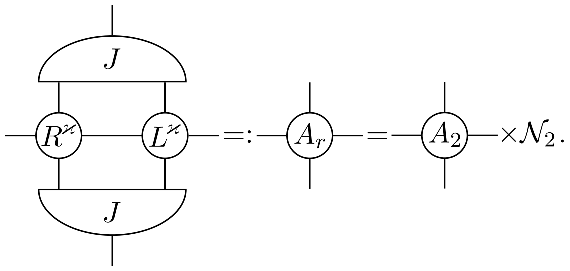
Proposition 2.6. *The functions $\widehat{b}_L$ and $\widehat{b}_R$ are monotonic and subhomogeneous on the (downward closed) sets on which they are defined. Functions b_L and b_R are analytic, $b_L(0) = b_R(0) = 0$.*

Proof. This proposition is analogous to Prop. 2.5, so we will be brief. Let us discuss why $\widehat{b}_L$ is subhomogeneous. Leaving aside $(\widehat{b}_L)_{\mathbf{oooo}} = 0$, the other elements of $\widehat{b}_L$ are obtained by combining in quadrature as in (2.78) tensor elements of $\widehat{L}$ which are all subhomogeneous, given by monomials in $\widehat{b}$ of order $\geqslant 1$, times scalar factors monotonic in $\widehat{b}$. Combining in quadrature preserves subhomogeneity. □

Recall that throughout Section 2.7, A was used as a shorthand for A_1, and b was used as a shorthand for b_1, where the end result of the gauge transformation was $A_1 = A_* + b_1$. When we compose the four steps of the 2x1 map, we will have to restore the subscript 1.

2.8 Reconnection and rotation

The last steps of our RG map is the reconnection followed by rotation, Fig. 1(c),(d). After disentangling and splitting, our tensor network consists of 2x1 blocks as in the r.h.s. of Fig. 1(b). Tensors L, R were defined in Section 2.7.5; at the end of which section we also replaced them with $L^{\varkappa}$, $R^{\varkappa}$. In terms of these latter tensors, reconnection is graphically described by the following diagram:



(2.83) ⟨code 25⟩

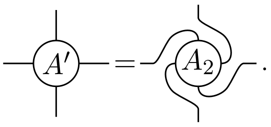
Here:

- J is an isometry from $V \times V$ onto V, defined below;
- A_r (r for reconnection) is defined as the shown contraction;
- Since A_r is not in general normalized, we define the normalized $A_2 = A_r/\mathcal{N}_2$, $\mathcal{N}_2 = (A_r)_{0000}$.

The reconnection step halves the horizontal size of the lattice. We have:

$$\boxed{Z(A_1, \ell_x \times \ell_y) = Z(A_r, \ell_x/2 \times \ell_y) = \mathcal{N}_2^{\mathrm{Vol}/2} Z(A_2, \ell_x/2 \times \ell_y)}\,. \tag{2.84}$$

After this, we perform the rotation step, Fig. 1(d), which we copy here:



(2.85)

This defines the final, normalized, tensor A' as the 90 degree counterclockwise rotation of A_2. So clearly

$$\boxed{Z(A_2, \ell_x/2 \times \ell_y) = Z(A', \ell_y \times \ell_x/2)}\,. \tag{2.86}$$

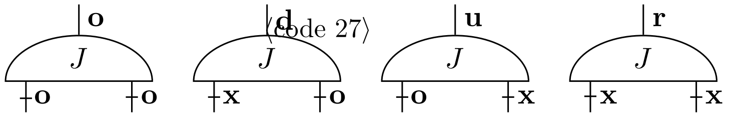
Combining Eqs. (2.44),(2.84),(2.86), we obtain (2.2) with $\mathcal{N} = \mathcal{N}_1^2\mathcal{N}_2$. We record this as:

Proposition 2.7. *The defined map $A \mapsto \mathcal{N}A'$ is an RG map (for any values of the reweighting parameters $w_{\mathbf{x}}, w_{\mathbf{o}} > 0$) in the sense that it preserves the tensor network partition function.*

Let us describe the isometry $J : V \otimes V \to V$. We split the "ingoing" vertical legs of J into sectors corresponding to $V = \mathbf{o} \oplus \mathbf{x}$. We split the "outgoing" leg into sectors $V = \mathbf{o} \oplus \mathbf{d} \oplus \mathbf{u} \oplus \mathbf{r}$. (This leg is vertical, but it will become horizontal after rotation, so we use 4 sectors as on the horizontal legs of A.) Then, we require that J maps isometrically the 4 sectors of $V \otimes V$ onto the 4 sectors of V as follows (see Fig. 7):

$$\mathbf{o} \otimes \mathbf{o} \to \mathbf{o}, \qquad \mathbf{o} \otimes \mathbf{x} \to \mathbf{u}, \qquad \mathbf{x} \otimes \mathbf{o} \to \mathbf{d}, \qquad \mathbf{x} \otimes \mathbf{x} \to \mathbf{r}. \tag{2.87}$$

The first three of these assignments and their rationale were already discussed in Section 2.3.4. The last assignment $\mathbf{x} \otimes \mathbf{x} \mapsto \mathbf{r}$ was not needed there, since the sector $\mathbf{x} \otimes \mathbf{x}$ does not arise in the linearized analysis. In the full nonlinear analysis, the sector $\mathbf{x} \otimes \mathbf{x}$ does arise and its information must be preserved; so we map it to another sector $\mathbf{r}$ foreseen specifically for this purpose.



⟨code 27⟩

Figure 7: Nonzero sectors of the J isometry. The "ingoing" legs are distinguished.

We next discuss the hat-tensors, starting with A_r. Note that the isometry J is not HS, so we use a check-tensor for it. The tensor J has two distinguished legs, indicated by tick marks in Fig. 7. The check-tensor $\check{J}$ has elements equal to 1 for the four elements corresponding to the four sectors in the figure. The rest of the elements are zero.

The contraction in Eq. (2.83) which defines A_r does not quite meet our definition of an allowed contraction (Definition 2.3) since the two distinguished legs are contracted with two different HS tensors. But there is an easy workaround by doing that contraction in two steps. We first contract $R^\varkappa$ with $L^\varkappa$, to get a HS tensor which is then contracted with J's. Both steps are allowed contractions and Lemma 2.2 applies. We thus see that the hat-tensor $\widehat{A}_r$ is given by replacing $R^\varkappa \to \widehat{R}^\varkappa$, $L^\varkappa \to \widehat{L}^\varkappa$, $J \to \check{J}$ in the l.h.s. of Eq. (2.83).

The hat-tensor for $A_2 = A_r/\mathcal{N}_2$ is now obtained similarly to $A_1 = A_g/\mathcal{N}_1$ in the gauge transformation discussion. We start by writing

$$\widehat{A}_r = \widehat{\mathcal{N}}_2 \widehat{A}_* + \widehat{b}_r, \qquad (\widehat{b}_r)_{\mathbf{oooo}} = 0. \qquad \langle\text{code 28}\rangle \tag{2.88}$$

Here $\widehat{\mathcal{N}}_2$, $\widehat{b}_r$ are functions of $\widehat{b}_L$ and $\widehat{b}_R$. $\widehat{\mathcal{N}}_2$ is an upper bound on $|\mathcal{N}_2|$, while a positive lower bound for $|\mathcal{N}_2|$ is given by $|\mathcal{N}_2|_{plb} = 2 - \widehat{\mathcal{N}}_2$, positive assuming $\widehat{\mathcal{N}}_2 < 2$. The hat-tensor for A_2 is then given by

$$\widehat{A}_2 = \widehat{A}_* + \widehat{b}_2, \qquad \widehat{b}_2 = \widehat{b}_r / |\mathcal{N}_2|_{plb}\,, \tag{2.89}$$

Finally, we obtain $\widehat{A}' = A_* + \widehat{b}'$ by rotating $\widehat{A}_2$ by 90 degrees counterclockwise.

We summarize the most important functions defined in this section:

$$\widehat{\mathcal{N}}_2 : \{(\widehat{b}_L, \widehat{b}_R) \in \widehat{\mathbb{H}}_0 \times \widehat{\mathbb{H}}_0\} \to \mathbb{R}_{\geqslant 0}, \tag{2.90}$$

$$\widehat{b}' : \widehat{\Omega}_2 \to \widehat{\mathbb{H}}_0, \quad \widehat{\Omega}_2 = \{(\widehat{b}_L, \widehat{b}_R) \in \widehat{\mathbb{H}}_0 \times \widehat{\mathbb{H}}_0 : \widehat{\mathcal{N}}_2 < 2\} \tag{2.91}$$

$$\mathcal{N}_2 : \{(b_L, b_R) \in \mathbb{H}_0 \times \mathbb{H}_0\} \to \mathbb{C}, \tag{2.92}$$

$$b' : \Omega_2 \to \mathbb{H}_0\,, \quad \Omega_2 = \{(b_L, b_R) \in \mathbb{H}_0 \times \mathbb{H}_0 : \exists\, \widehat{b}_L, \widehat{b}_R \text{ for } b_L, b_R \text{ such that } (\widehat{b}_L, \widehat{b}_R) \in \widehat{\Omega}_2\}\,. \tag{2.93}$$

The following proposition is analogous to Props. 2.5,2.6; we omit the proof.

Proposition 2.8. *$\widehat{\mathcal{N}}_2$ is monotonic. $\widehat{b}'$ is monotonic and subhomogeneous on the (downward closed) set $\widehat{\Omega}_2$. Functions $\mathcal{N}_2$, b' are analytic. We have $b'(0,0) = 0$, $\mathcal{N}_2(b_L, b_R) = 1 + O(b_L b_R)$, $\mathcal{N}_2 \neq 0$ in Ω_2, and the bound:*

$$|\mathcal{N}_2 - 1| \leqslant \widehat{\mathcal{N}}_2 - 1\,. \tag{2.94}$$

2.9 Master function

All the steps of the 2x1 map have now been defined. Composing them, we define the map $b'(b)$ on the set Ω as in (2.28) where the associated set $\widehat{\Omega}$ is given by:

$$\begin{aligned}\widehat{\Omega} = \{\widehat{b} \in \widehat{\mathbb{H}}_0 :\ & \widehat{\beta}_h(\widehat{b}) < 1,\ \widehat{\beta}_v(\widehat{b}) < 1,\ \widehat{\mathcal{N}}_1(\widehat{b}) < 2,\\ & \widehat{\alpha}(\widehat{b}_1) < 1 \text{ where } \widehat{b}_1 = \widehat{b}_1(\widehat{b}),\\ & \widehat{\mathcal{N}}_2(\widehat{b}_L, \widehat{b}_R) < 2 \text{ where } \widehat{b}_L = \widehat{b}_L(\widehat{b}_1), \widehat{b}_R = \widehat{b}_R(\widehat{b}_1)\}.\end{aligned} \tag{2.95}$$

The associated hat-tensor map $\widehat{b}'(\widehat{b}) =: \mathfrak{M}(\widehat{b})$ is the master function, defined on $\widehat{\Omega}$.

The $\mathcal{N}$-factor is defined on Ω by

$$\mathcal{N}(b) = (\mathcal{N}_1(b))^2 \mathcal{N}_2(b_L, b_R) \text{ where } b_L = b_L(b_1(b)), b_R = b_R(b_1(b))\,. \tag{2.96}$$

The following proposition is obtained by putting together Props. 2.5,2.6,2.8; we omit the proof.

Proposition 2.9. *The master function $\mathfrak{M}(\widehat{b})$ is monotonic and subhomogeneous on the downward closed set $\widehat{\Omega}$. Functions $\mathcal{N}(b)$, $b'(b)$ are analytic on Ω. We have $b'(0) = 0$, $\mathcal{N}(b) = 1 + O(b^2)$, $\mathcal{N}(b) \neq 0$ in Ω.*

2.10 Constraints on $w_{\mathbf{x}}$ and $w_{\mathbf{o}}$ from linear stability

Now that the 2x1 map has been described in detail, we would like to come back to the question about the numerical values of parameters $w_{\mathbf{x}}$ and $w_{\mathbf{o}}$. We would like to require that the linearization of the RG map is a contraction in some norm, as discussed in Section 2.3. Let us use the leading part of the master function to control the linearization of the RG map. We write

$$b'(b) = b'^{(1)} + O(b^2) \tag{2.97}$$

where $b'^{(1)}$ is linear in b. We can get a hat-tensor for $b'^{(1)}$ by extracting the leading part of the master function, as follows:

$$\widehat{b}'^{(1)} = \lim_{\epsilon \to 0} \epsilon^{-1} \mathfrak{M}(\epsilon \widehat{b})\,. \tag{2.98}$$

Our computer code for the master function $\mathfrak{M}(\widehat{b})$ [8] (see also App. A) is suitable for both numerical and symbolic calculations. Numerical mode will be used in Section 3 for controlling the map at the full nonlinear level. Using the symbolic mode of the code, we easily evaluate (2.98) and obtain [8]:

$$\begin{aligned}
\widehat{b}'^{(1)}_{\mathbf{uxoo}} &= \left[W\,\widehat{b}^2_{\mathbf{oxuo}} + (1 + 1/w_{\mathbf{x}}^2)\,\widehat{b}^2_{\mathbf{oxro}} + (1 + w_{\mathbf{o}}^2 + 1/w_{\mathbf{x}}^2)\,\widehat{b}^2_{\mathbf{oxdo}}\right]^{1/2},\\
\widehat{b}'^{(1)}_{\mathbf{oxuo}} &= \left[W\,\widehat{b}^2_{\mathbf{oodx}} + (1 + 1/w_{\mathbf{x}}^2)\,\widehat{b}^2_{\mathbf{oorx}} + (1 + w_{\mathbf{o}}^2 + 1/w_{\mathbf{x}}^2)\,\widehat{b}^2_{\mathbf{ooux}}\right]^{1/2},\\
\widehat{b}'^{(1)}_{\mathbf{doox}} &= \left[W\,\widehat{b}^2_{\mathbf{uxoo}} + (1 + 1/w_{\mathbf{x}}^2)\,\widehat{b}^2_{\mathbf{rxoo}} + (1 + w_{\mathbf{o}}^2 + 1/w_{\mathbf{x}}^2)\,\widehat{b}^2_{\mathbf{dxoo}}\right]^{1/2},\\
\widehat{b}'^{(1)}_{\mathbf{oodx}} &= \left[W\,\widehat{b}^2_{\mathbf{doox}} + (1 + 1/w_{\mathbf{x}}^2)\,\widehat{b}^2_{\mathbf{roox}} + (1 + w_{\mathbf{o}}^2 + 1/w_{\mathbf{x}}^2)\,\widehat{b}^2_{\mathbf{uoox}}\right]^{1/2},\\
\widehat{b}'^{(1)}_{\mathbf{uxuo}} &= \left[(1 + w_{\mathbf{o}}^2 + 1/w_{\mathbf{x}}^2)\,\widehat{b}^2_{\mathbf{oxdx}} + (1 + w_{\mathbf{o}}^2 + 1/w_{\mathbf{x}}^2)\,\widehat{b}^2_{\mathbf{oxux}} + (1 + 1/w_{\mathbf{x}}^2)\,\widehat{b}^2_{\mathbf{oxrx}}\right]^{1/2},\\
\widehat{b}'^{(1)}_{\mathbf{dodx}} &= \left[(1 + w_{\mathbf{o}}^2 + 1/w_{\mathbf{x}}^2)\,\widehat{b}^2_{\mathbf{dxox}} + (1 + w_{\mathbf{o}}^2 + 1/w_{\mathbf{x}}^2)\,\widehat{b}^2_{\mathbf{uxox}} + (1 + 1/w_{\mathbf{x}}^2)\,\widehat{b}^2_{\mathbf{rxox}}\right]^{1/2},\\
\widehat{b}'^{(1)}_{\mathbf{dodo}} &= \widehat{b}_{\mathbf{oxox}},\\
\widehat{b}'^{(1)}_{\mathbf{uouo}} &= \widehat{b}_{\mathbf{oxox}},
\end{aligned} \tag{2.99}$$

where we defined

$$W = 2\left(1/w_{\mathbf{x}}^2 + 1/w_{\mathbf{o}}^2\right). \tag{2.100}$$

Square roots in (2.99) are a consequence of combining in quadrature in (2.78), which makes $\mathfrak{M}(\widehat{b})$ non-analytic in $\widehat{b}$, see Remark 2.6. Checking (2.99) by pencil and paper would be a straightforward but tedious task; we haven't done it.

The special sectors of $\widehat{b}'^{(1)}$ and of $\widehat{b}$ are shown in (2.99) in red. (The special sectors were defined in Eq. (2.17).) Only special sectors appear in the l.h.s. This means that the non-special sectors of $b'^{(1)}$ vanish. This is as discussed in Section 2.3.

The non-special sectors do appear in the r.h.s. of (2.99). But as discussed in Section 2.3, we may focus on the special-to-special part of the linearization, denoted there K_{ss}, and it's enough to ascertain that the operator norm (with respect to the HS norm of tensors) $\|K_{ss}\| < 1$. Using (2.99), we have the bound

$$\|K_{ss}\| \leqslant \sqrt{W}\,. \tag{2.101}$$

Therefore, for $W < 1$ our map will be a contraction at the linearized level in the $|||\cdot|||$ norm described in Section 2.3. Since the 2x1 map is analytic, it will also remain a contraction at the full nonlinear level in a sufficiently small neighborhood of the origin and in the same norm. So, once RG iterations get sufficiently close to the origin, further iterations will decrease exponentially fast.

Below we will always choose parameters $w_{\mathbf{x}}$ and $w_{\mathbf{o}}$ so that $W < 1$. However we will no longer discuss the norm in which the map is a contraction. Instead, we will switch to controlling the map through the master function, as described in Section 2.5. This method can efficiently control both the early RG iterations, until the exponential decay sets in, and, thanks to Key Lemma 2.3, later exponentially decaying iterations, saving the effort to exhibit a contracting norm.

2.11 Symmetries

In conclusion, let us discuss symmetries of the 2x1 RG map, starting with the lattice symmetries. The square lattice has $\mathbb{Z}_4$ discrete rotation symmetry and $\mathbb{Z}_2 \times \mathbb{Z}_2$ reflection symmetries in the horizontal and vertical direction. The 2x1 map does not preserve these symmetries. For the rotation symmetry this is quite obvious, since the horizontal and the vertical directions are treated asymmetrically.

As for the reflection symmetries, they are violated by the assignment of minus signs for the ρ_4, ρ_5 tensors relative to the λ_4, λ_5 tensors (Section 2.7.1). Notice however that when we defined the master function, we replaced all minus signs by plus signs. So the master function preserves the reflection symmetries. (Note that the vertical reflection exchanges $\mathbf{u}$ and $\mathbf{d}$.)

Let us discuss next global symmetries. For lattice spin models, a global symmetry acts on spin values in a way which preserves the energy of a spin configuration. When such a model is translated into a tensor network, the global symmetry is inherited by the resulting tensor. The tensor network partition function remains invariant when we act on the tensor by gauge transformations, Fig. 1(a). The global symmetry group is formed by gauge transformations which leave the tensor invariant. In practice, consequences of a global symmetry group are best understood by decomposing the Hilbert spaces of vertical and horizontal legs into irreducible representations. Global symmetry then manifests itself through selection rules: a tensor element can be nonzero only if the tensor product of the four respective representations contains the trivial representation. We will see some examples in Section 3.

In our description of the 2x1 map we did not keep track of the possible global symmetry $\mathcal{G}$. It is possible to generalize the discussion. It is natural to assume that index 0 transforms in the trivial representation. Then, one can show that the 2x1 map preserves the global symmetry. Namely, all tensors which entered the definition of the map can be consistently defined so that they are $\mathcal{G}$-invariant, provided that b is $\mathcal{G}$-invariant. In particular the final tensor b' will be $\mathcal{G}$-invariant. To take advantage of this, it would be necessary to define a more fine-grained master function, dividing various sectors into subsectors corresponding to different irreps. This will not be done in this work.

3 Computer-assisted bounds on the high-T phase

In this section, we will use the 2x1 map defined in Section 2 to prove results in Table 1 about the high-T phase of lattice models. We consider general tensors first, and a couple of specific models later.

3.1 General tensors

Let $\mathbb{O}$ be a set containing the origin in $\mathbb{H}_0$ (which may or may not be open). Consider a tensor network built out of tensor $A = A_* + b$, $b \in \mathbb{O}$. Fix some values of reweighting parameters of the 2x1 map. We will say that $\mathbb{O}$ is a *basin of stability* if for any $b = b^{(0)} \in \mathbb{O}$:

- the 2x1 map can be iterated infinitely many times;
- the resulting sequence of tensors $b^{(i)}$, $i = 1, 2, \ldots$, tends to zero in the HS norm.

For $\delta > 0$, we consider a closed neighborhood of the origin in $\mathbb{H}_0$ given by:

$$\mathbb{O}_\delta = \{b \in \mathbb{H}_0 : \|b_{abcd}\| \leqslant \delta \text{ for all } abcd \neq \mathbf{oooo}\}\,. \tag{3.1}$$

The following theorem is our first main result.

Theorem 3.1. *(Stability of the high-T fixed point) Fix the reweighting parameters $w_{\mathbf{x}} = 2.2$, $w_{\mathbf{o}} = 2$. Then $\mathbb{O}_\delta$, $\delta = 0.02$, is a basin of stability.*

Proof. To control the 2x1 map we use the master function as described in Section 2.5. We define a hat-tensor $\widehat{b}^{(0)} \in \widehat{\mathbb{H}}_0$ with $b^{(0)}_{\mathbf{oooo}} = 0$ and all other components $\widehat{b}^{(0)}_{abcd} = \delta$. We then start acting on $\widehat{b}^{(0)}$ repeatedly with the master function $\mathfrak{M}$, generating a sequence of hat-tensors $\widehat{b}^{(i)}$, $i = 1, 2, \ldots$. The idea is that $\widehat{b}^{(0)}$ defines a box around A_* which flows under RG, and we need to arrange that it flows to zero size.

We perform the first $i_0 + 1$ iterations numerically using our computer code. At every step the code checks that the master function is defined and if so, it evaluates the next $\widehat{b}^{(i)}$. Suppose that all these evaluations went through, and the last two tensors satisfy (2.30) which we copy here:

$$\widehat{b}^{(i_0+1)} \leqslant \lambda\, \widehat{b}^{(i_0)} \text{ for some } \lambda < 1. \tag{3.2}$$

For $\delta = 0.02$, $w_{\mathbf{x}} = 2.2, w_{\mathbf{o}} = 2$, our calculations [8] show that this holds for $i_0 = 15$ with $\lambda = 0.96$. (Our code evaluates $\widehat{b}^{(i)}$ and checks (3.2) using interval arithmetic so that the argument is rigorous.) Given (3.2), Key Lemma 2.3 tells us that all further iterations of the master functions are also defined and tend to zero.

By Proposition 2.9, since the master functions iterates are defined, the 2x1 map iterates $b^{(i)}$, $i = 1, 2, \ldots$ are also defined for any $b^{(0)}$ for which $\widehat{b}^{(0)}$ is a hat-tensor, i.e. for any $b^{(0)} \in \mathbb{O}_\delta$, and moreover $\widehat{b}^{(i)}$ is a hat tensor for $b^{(i)}$. Therefore, $b^{(i)}$ tends to zero in the HS norm. □

How did we arrive at the above numerical values of $\delta, w_{\mathbf{x}}, w_{\mathbf{o}}$? We took $w_{\mathbf{x}}$, $w_{\mathbf{o}}$ within the range $W < 1$ but not too large, as the latter would unnecessarily enhance K_{ns} and the nonlinear terms. We ran master function iterations for several values of $w_{\mathbf{x}}$ and $w_{\mathbf{o}}$ near 2, and for several values of δ, producing plots like Fig. 8 below. We then picked $w_{\mathbf{x}} = 2.2$ and $w_{\mathbf{o}} = 2$ as giving us the largest δ among those we tried, for which we could see exponential decrease set in after 10-20 iterations. Of course, varying $w_{\mathbf{x}}$ and $w_{\mathbf{o}}$ a bit, the theorem still holds for a somewhat different δ. In the future, it would be interesting to optimize them systematically and to establish an even larger basin of attraction of the high-T fixed point; see Section 4 for a discussion.

In Fig. 8 we show the sequence of hat-tensors as a function of the RG step i, for $\delta, w_{\mathbf{x}}, w_{\mathbf{o}}$ as in the theorem. Note that many components of $\widehat{b}^{(i)}$ behave nonmonotonically in the initial iterations, but eventually the exponential decay sets in, and condition (3.2) can be verified. The initial nonmontonic behavior suggests that it would be challenging, or perhaps impossible, to prove Theorem 3.1 by exhibiting a norm in which the 2x1 map is a contraction in the whole of $\mathbb{O}_\delta$.

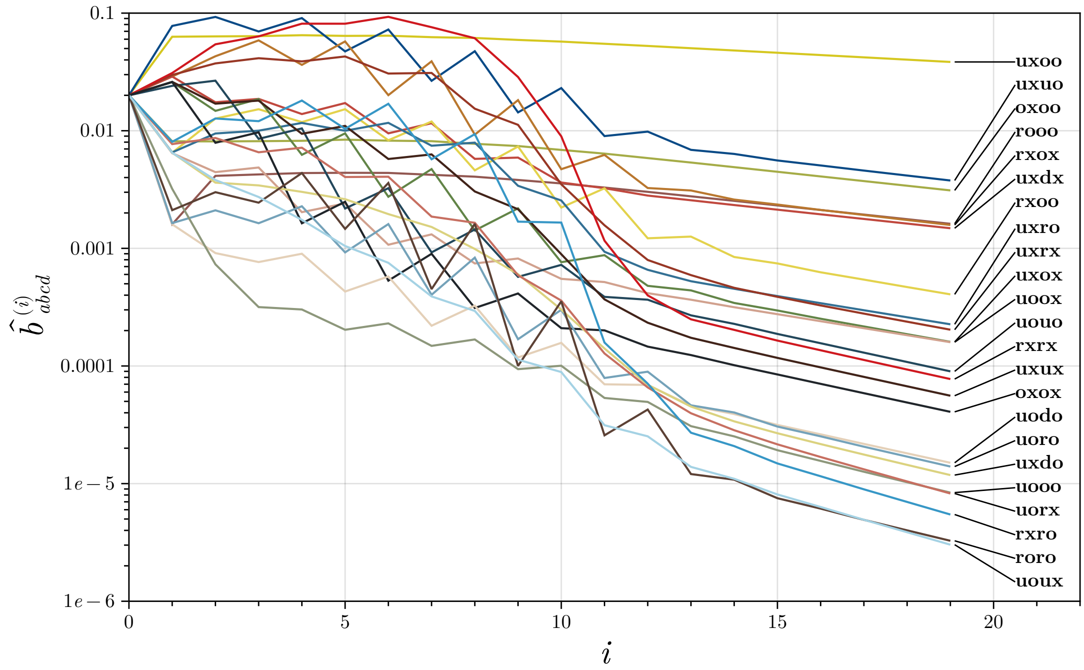


Figure 8: Master function iterates with the initial conditions $\widehat{b}^{(0)}_{\mathbf{oooo}} = 0$, $\widehat{b}^{(0)}_{abcd} = \delta = 0.02$ for all other components, and for the reweighting parameters $w_{\mathbf{x}} = 2.2$, $w_{\mathbf{o}} = 2$. We only plot one sector per each group related by the lattice reflections. This is because the reflection symmetry of the initial $\widehat{b}^{(0)}$ is preserved by the master function, see Section 2.11. (Our interval arithmetic calculations are initialized with precision 10^{-21}. After 20 iterations the precision is 10^{-17}, i.e. completely negligible on the scale of this plot.)

We next discuss the free energy. The tensor network free energy per site is defined by [4][19]

$$f(A) = \lim_{\ell\to\infty} f(A, \ell\times\ell), \quad f(A, \ell_x \times \ell_y) = \frac{1}{\ell_x \ell_y} \log Z(A, \ell_x \times \ell_y)\,. \tag{3.3}$$

We will write $f(b)$ to denote $f(A_* + b)$ hoping that this will not lead to a confusion. For Hamiltonians with finite-range interactions, and hence for tensor networks obtained by translating partition functions of such Hamiltonians, the existence of such a limit is a classic result [11]. For an arbitrary A the limit may not always exist.

However, if the RG iterates tend to zero, then it's easy to show that the limit exists along a subsequence and moreover the free energy is analytic. Thus all such tensors belong to the same phase, which we call the high-T phase. More precisely we have the following result:

Proposition 3.2. *Suppose $\mathbb{O}$ is a basin of stability. Then for any $b \in \mathbb{O}$ the limit defining the free energy exists along the subsequence $\ell = 2^k$. The free energy is given by the sum of the following series:*

$$f(b) = \sum_{i=0}^{\infty} 2^{-i-1} \log \mathcal{N}(b^{(i)})\,. \tag{3.4}$$

[19] Sometimes the free energy is defined with the opposite sign.

It is analytic in the interior of $\mathbb{O}$.

Proof. This is fully analogous to [4], Prop. 4.3, so we will be brief. (The RG map there had scale factor 4.) By Prop. 2.9, $\mathcal{N}(b)$ is analytic and nonzero in Ω, and $\mathcal{N}(b) = 1 + O(b^2)$. Since $b^{(i)} \to 0$, the series in (3.4) converges, and its sum is analytic in the interior of $\mathbb{O}$. Furthermore, by repeatedly using (2.2), we show that $f(b, 2^{k+1} \times 2^{k+1})$ equals the partial sum of (3.4) over $i \leqslant 2k+1$ plus the "remainder" $2^{-2k} f(b^{(2k)}, 2 \times 2)$. As $k \to \infty$, the remainder tends to zero, and we are done. □

Combining this with Theorem 3.1, we get

Proposition 3.3. *The free energy is analytic in the interior of $\mathbb{O}_\delta$, $\delta = 0.02$.*

It's interesting to discuss how these results compare to the cluster expansion. The cluster expansion [12] also provides an analytic representation for the free energy in the region where it converges. The cluster expansion computes the Taylor expansion of the free energy of a lattice model in polymer activities. As discussed in [3, App. B], the free energy of a tensor network can also be computed via a cluster expansion, with polymers being connected subsets of k lattice points, polymer activity bounded by $\|b\|^k$, and the hardcore repulsion as the compatibility condition. Applying a standard criteria of cluster expansion convergence such as Kotecký-Preiss [13], our back of the envelope estimate gives convergence for $\|b\| < 7 \times 10^{-4}$.[20] We don't want to claim that this is the best the cluster expansion can do. But it does look like our RG method is already doing better than the cluster expansion. Furthermore it can be significantly improved, as we will discuss in Section 4.

Unlike the cluster expansion of the free energy, Eq. (3.4) is not a Taylor expansion. Term number i in this equation, if expanded in b, would contain terms of all orders. Eq. (3.4) converges extremely rapidly for two reasons. First, because of the factor 2^{-i-1} and second because $b^{(i)} \to 0$ and $\mathcal{N} = 1 + O(b^2)$. For any concrete $b^{(0)}$, one can take advantage of this fast convergence by evaluating a few terms of the expansion explicitly and estimating the rest.

For generic $b^{(0)}$, we can get the following bounds on the free energy in $\mathbb{O}_\delta$. Let us write $\widehat{\mathcal{N}}_1 = 1 + \widehat{n}_1$, $\widehat{\mathcal{N}}_2 = 1 + \widehat{n}_2$. Then, using $\mathcal{N} = \mathcal{N}_1^2 \mathcal{N}_2$ we have the lower and upper bounds

$$\begin{aligned} C_i^- \leqslant \log \mathcal{N}(b^{(i)}) \leqslant C_i^+ , \\ C_i^\pm = 2 \log[1 \pm \widehat{n}_1(\widehat{b}^{(i)})] + \log[1 \pm \widehat{n}_2(\widehat{b}^{(i)})] \,. \end{aligned} \qquad \langle\text{code } 37\rangle\, (3.5)$$

Note that C_i^- is negative while C_i^+ is positive.

We can use (3.5) to bound every term in the series (3.4). We split the resulting series into the head $i \leqslant i_0$ plus the tail $i > i_0$, where i_0 is as in the proof of Theorem 3.1. The head can be evaluated since $\widehat{n}_1(\widehat{b}^{(i)})$, $\widehat{n}_2(\widehat{b}^{(i)})$ are available in the process of iterating the master function. A tail term number i can be estimated by $2^{i_0 - i}$ times the last head term, since $\widehat{n}_1$ and $\widehat{n}_2$ are monotonic, and we know from Key Lemma 2.3 that $\widehat{b}^{(i)} \leqslant \widehat{b}^{(i_0)}$. We thus obtain free energy bounds valid for any $b \in \mathbb{O}_\delta$:

$$\sum_{i=0}^{i_0} (1 + \delta_{i,i_0}) 2^{-i-1} C_i^- \leqslant f(b) \leqslant \sum_{i=0}^{i_0} (1 + \delta_{i,i_0}) 2^{-i-1} C_i^+ \,, \qquad (3.6) \qquad \langle\text{cod}$$

where δ_{i,i_0} is the Kronecker delta, $\delta_{i,i_0} = 1$ for $i = i_0$ and is zero otherwise.

Bound (3.6) may appear not particularly impressive, for example it does not determine the sign of $f(b)$. We show it to demonstrate the general idea. As mentioned above, in a more meaningful treatment we need to evaluate the first few terms of (3.4) explicitly, and use a bound like (3.6) on the remaining terms. This is expected to give an extremely accurate rigorous estimate of free energy throughout the basin of stability.

[20] This cluster expansion is similar in structure to the Ising model in strong magnetic field discussed in [12, Section 5.7.1], replacing their polymer activities by the $\|b\|^k$ bound, $k = |S|$. Note that $k \geqslant 2$ in our case. We choose $a(S) = |[S]_1|$ as in [12], and use their Exercise 5.3 to bound the number of polymers.

3.2 Ising model

Let us examine next how well our method can detect the high-T phase in some familiar lattice models, starting with the (nearest-neighbor) Ising model defined by the partition function:

$$Z = \sum_{\sigma_x=\pm 1} \exp\Big(\beta \sum_{\langle xy\rangle} \sigma_x\sigma_y\Big), \tag{3.7}$$

where x are vertices of the square lattice and $\sum_{\langle xy\rangle}$ means summation over nearest-neighbor pairs. The Ising model can be represented as a tensor network composed of the tensor $A = A(\beta)$ with bond dimension 2 and the following nonzero tensor elements [14],[3, App. A.1]

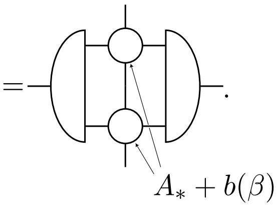
$$A_{0000} = \cosh(4\beta) + 3, \quad A_{0101} = A_{1111} = \cosh(4\beta) - 1, \quad A_{0011} = \sinh(4\beta), \tag{3.8}$$

and rotations thereof.[21] The $\mathbb{Z}_2$ symmetry of the Ising model is reflected in A as follows (see Section 2.11). The state 0 is $\mathbb{Z}_2$-even i.e. transforms in the trivial representation. The state 1 is $\mathbb{Z}_2$-odd i.e. transforms in the nontrivial representation. All elements of A with an odd number of 1 indices are zero. However, we will not take advantage of the $\mathbb{Z}_2$ symmetry in the analysis below.

As usual we normalize the tensor by writing $A = A_{0000}(A_* + b(\beta))$ with $b(\beta) \in \mathbb{H}_0$. Note that $b(\beta) = O(\beta)$ as $\beta \to 0$ and has nonnegative tensor elements which are monotonic functions of β.

We would like to find $\bar\beta$ such that $A_* + b(\beta)$ for $\beta \leqslant \bar\beta$ flows to A_* under tensor RG. This will imply that the Ising model for $\beta \leqslant \bar\beta$ is in the high-T phase. Already Theorem 3.1 could be used to get an estimate for $\bar\beta$, but we will instead get a much better result using the explicit form of A.

We find it natural to do a preliminary RG step. We define tensor $A^{(0)}$ (which will be the initial tensor for 2x1 map iterations) as:

$A^{(0)} :=$  . (3.9)

I.e. $A^{(0)}$ is a contraction of two copies of $A_* + b$ put on top of each other, followed by an isometry on horizontal legs. We choose this isometry so that it maps 00 to 0,[22] 10 to 1, 01 to 2, 11 to 3.

We think of $A^{(0)}$ having infinitely many indices, but only a finite number are "active" i.e. give rise to nonzero components: 2 vertical and 4 horizontal. We define one horizontal and vertical sectors for $A^{(0)}$ per each active index (distributing inactive indices into **x**, **u**, **d**, **r** so that each of these sectors is infinite-dimensional):

$$\begin{aligned} &\text{vertical:} && \mathbf{o} = \operatorname{span}(e_0), \quad \mathbf{x} = \operatorname{span}(e_1, \ldots), \\ &\text{horizontal:} && \mathbf{o} = \operatorname{span}(e_0), \quad \mathbf{u} = \operatorname{span}(e_1, \ldots), \quad \mathbf{d} = \operatorname{span}(e_2, \ldots), \quad \mathbf{r} = \operatorname{span}(e_3, \ldots). \end{aligned} \tag{3.10}$$

We see that $A^{(0)}$, unlike A, allows a natural sector decomposition $\mathbf{o} \oplus \mathbf{d} \oplus \mathbf{u} \oplus \mathbf{r}$ on the horizontal legs, which is the reason why we defined it.

$A^{(0)}$ is a normalized tensor. We write $A^{(0)} = A_* + b^{(0)}(\beta)$ with $b^{(0)}(\beta) \equiv b^{(0)}_{\text{Ising}}(\beta) \in \mathbb{H}_0$. (We will use the "Ising" subscript only in formulations of main results.)

The tensor $b^{(0)}(\beta)$ is $O(\beta)$ for $\beta \to 0$ and has nonnegative tensor elements which are monotonic functions of β. It inherits these properties from $b(\beta)$.

[21] To be precise, tensor network partition function $Z(A, \ell \times \ell)$ with periodic boundary conditions equals the Ising model partition functions on the square lattice rotated by 45 degrees, which contains $2\ell^2$ spins. This needs to be kept in mind when comparing the free energy per site.

[22] This is a shorthand for $e_0 \otimes e_0 \mapsto e_0$.

Let $\widehat{b}^{(0)}(\beta)$ be the minimal hat-tensor for $b^{(0)}(\beta)$. Because $b^{(0)}(\beta)$ is a nonnegative tensor and because every sector contains one active index, $\widehat{b}^{(0)}(\beta)$ is obtained from $b^{(0)}(\beta)$ replacing active indices by sector labels.

Instead of the "square" neighborhood $\mathbb{O}_\delta$ in Theorem 3.1, we will need more general "rectangular" subsets of $\mathbb{H}_0$ defined by

$$\mathbb{O}(\widehat{b}) = \{b \in \mathbb{H}_0 : \|b_{abcd}\| \leqslant \widehat{b}_{abcd} \text{ for all } abcd \neq \mathbf{oooo}\}\,, \tag{3.11}$$

where $\widehat{b}$ is any hat-tensor with $\widehat{b}_{\mathbf{oooo}} = 0$. In other words, $\mathbb{O}(\widehat{b})$ is a set of all $b \in \mathbb{H}_0$ for which $\widehat{b}$ is a hat-tensor. If all non-**oooo** components of $\widehat{b}$ are positive, $\mathbb{O}(\widehat{b})$ is a closed neighborhood of the origin, otherwise it has empty interior.

The proof of Theorem 3.1 starts by considering the tensor $\widehat{b}^{(0)}$ whose non-**oooo** components are all equal to δ. We now repeat that proof with a different $\widehat{b}^{(0)}$, namely $\widehat{b}^{(0)} = \widehat{b}^{(0)}(\bar{\beta})$ defined as above. This tensor defines a rectangular box around A_*, whose size grows with $\bar{\beta}$. This box then flows under RG. We would like to take this initial box as large as possible, by increasing $\bar{\beta}$, while preserving the condition that it eventually flows to zero size. We ensure the latter by checking that master functions iterates initialized at $\widehat{b}^{(0)}(\bar{\beta})$ start converging exponentially after 10-20 steps. We can also tweak the reweighting parameters $w_{\mathbf{x}}$ and $w_{\mathbf{o}}$.

Proposition 3.4. *(a) The set of tensors $\mathbb{O}(\widehat{b}^{(0)}_{\text{Ising}}(\bar{\beta}))$, $\bar{\beta} = 0.12$, is a basin of stability for the 2x1 RG map with the reweighting parameters $w_{\mathbf{x}} = 2.3$, $w_{\mathbf{o}} = 2$.*

(b) The free energy is analytic on a neighborhood of $\mathbb{O}(\widehat{b}^{(0)}_{\text{Ising}}(\bar{\beta}))$.

(c) $b^{(0)}_{\text{Ising}}(\beta) \in \mathbb{O}(\widehat{b}^{(0)}_{\text{Ising}}(\bar{\beta}))$ for any $0 \leqslant \beta \leqslant \bar{\beta}$.

Proof. Part (a) is proved by the strategy from the preceding paragraph. Fig. 9 shows the corresponding master function iterates initialized at $\widehat{b}^{(0)}_{\text{Ising}}(\bar{\beta})$. Condition (3.2) is satisfied for $i_0 = 15$ with $\lambda = 0.95$ [8]. This proves part (a) by the arguments analogous to Theorem 3.1.

We next argue for (b). Note that $\mathbb{O}(\widehat{b}^{(0)}(\beta))$ has empty interior, some non-**oooo** elements of $\widehat{b}^{(0)}(\beta)$ being zero, although starting from $\widehat{b}^{(1)}(\beta)$ they are all nonzero (see Fig. 9).[23] So we cannot appeal directly to Proposition 3.2, but there is an easy workaround. Let $\widehat{b}^{(0)}(\bar{\beta}, \epsilon)$ be a regularization of $\widehat{b}^{(0)}(\bar{\beta})$ where all non-**oooo** elements were increased by an $\epsilon > 0$. When ϵ is very small, the first i_0 master function iterates of $\widehat{b}^{(0)}(\bar{\beta}, \epsilon)$ are very close to those of $\widehat{b}^{(0)}(\bar{\beta})$. The given proof of Proposition 3.4 then implies that $\mathbb{O}(\widehat{b}^{(0)}(\bar{\beta}, \epsilon))$ is also a basin of stability, for a sufficiently small $\epsilon > 0$. The latter set has nonempty interior, which contains $\mathbb{O}(\widehat{b}^{(0)}(\bar{\beta}))$, and on which free energy is analytic by Proposition 3.2.

Finally, (c) follows from the fact, stated above, that $\widehat{b}^{(0)}_{\text{Ising}}(\beta)$ is monotonic in β. □

Let us recap. We have translated the Ising model into a tensor network. We have shown that in the range $\beta \leqslant \bar{\beta} = 0.12$ the free energy of the resulting tensor network is analytic under arbitrary perturbations of sufficiently small HS norm (which include perturbations which break $\mathbb{Z}_2$ symmetry such as the magnetic field). This proves that the Ising model for $\beta \leqslant 0.12$ is in the high-T phase. Note that the Ising model critical point is located at $\beta_c = \frac{1}{2}\log\big(1 + \sqrt{2}\big) \approx 0.44$. Thus, our RG argument captured roughly one quarter of the full high-T phase $\beta < \beta_c$.

For the sake of the argument, let us compare to the cluster expansion. One published estimate that we found is $e^\beta \beta < 1/192$ [15, Section V.7, Example 2]. Using the Kotecký-Preiss condition [13] we get a slightly better result $\beta < 0.009$. This can certainly be optimized further, but so can our approach, see Section 4.

[23] Let's explain why this happens, taking sector **uxuo** as an example. Since we are not subdividing the infinite-dimensional sectors by symmetry, there are $\mathbb{Z}_2$-even indices in **u**, **x**. For the initial tensor $b^{(0)}$, those indices are inactive, but they become active at subsequent iterations. See also the discussion in Section 2.11.

⟨code 41⟩

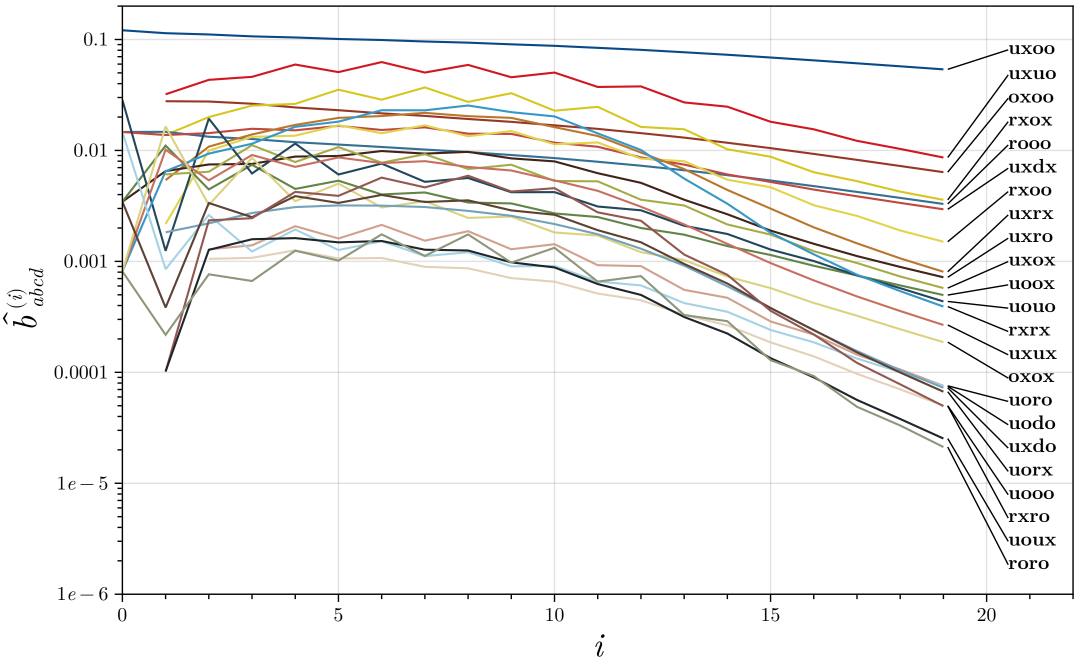


Figure 9: Master function iterates initialized at the tensor $\widehat{b}^{(0)}_{\text{Ising}}(\bar{\beta})$, $\bar{\beta} = 0.12$. The reweighting parameters are: $w_{\mathbf{o}} = 2.0$ and $w_{\mathbf{x}} = 2.3$. Some of the trajectories appear only starting from the second step, as the initial tensor has these sectors identically zero.

To conclude this section, let us use the master function iterates to get estimates of the free energy. The free energy per site of the original tensor network obtained by translating the Ising model is given by

$$f(A(\beta)) = \log(\cosh(4\beta) + 3) + \frac{1}{2} f(b^{(0)}(\beta))\,, \tag{3.12}$$

where the first term comes from the normalization of A and the factor $1/2$ in front of the second term is because the preliminary RG step halved the number of tensors in the network. The free energy $f(b^{(0)})$ is computed by the series (3.4). As discussed below Proposition 3.3, for a proper treatment we should evaluate the first few terms in (3.4) and estimate the rest. Since our purpose here is just to demonstrate the general idea, we will be content to simply bound all terms in $f(b^{(0)}(\beta))$ using the general formula (3.6). Note that since $b^{(0)}(\beta)$ is a nonnegative tensor, the free energy is non-negative, and only the upper bound (3.6) is of interest. So we have:

$$0 \leqslant f(b^{(0)}(\beta)) \leqslant \sum_{i=0}^{i_0} (1 + \delta_{i,i_0}) 2^{-i-1} C_i^+\,, \tag{3.13}$$

where C_i^+ are given by Eq. (3.5). The bound on $f(b^{(0)}(\beta))$ and the resulting bounds on $f(A(\beta))$ are shown in Fig. 10 for $\beta \leqslant 0.12$. In the same plots we include the exact value of these quantities, extracted using Onsager's exact solution (see footnote 21). For $f(A(\beta))$, the dominant contribution comes from the first term in (3.12). So the uncertainty due to $f(b^{(0)}(\beta))$ is hardly visible in much of the range of β.

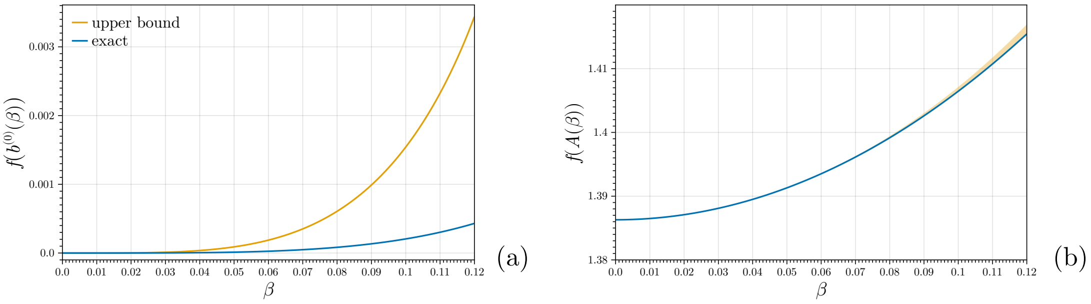


Figure 10: Bounds on the Ising model free energy compared to the exact values. (a) Orange: Upper bound (3.13) on $f(b^{(0)}(\beta))$, computed for $\beta = 0 \dots 0.12$ with step 10^{-3}. For each β, master function iterations were performed up to $i_0 = 15$, checking the Key Lemma condition at the end. Blue: the exact value. (b) Shaded region: the allowed range of $f(A(\beta))$. Blue line: the exact value from Onsager's solution.

3.3 XY model

Our last example is the (nearest-neighbor) XY model, defined by the following partition function:

$$Z = \int \prod_x \frac{d\theta_x}{2\pi} \exp\Big(\beta \sum_{\langle xy \rangle} \cos(\theta_x - \theta_y)\Big), \tag{3.14}$$

where x are vertices of the square lattice, and the variables θ_x live on the unit circle $S^1 \simeq [0, 2\pi)$. Since the space of spin states is continuous, tensor network representation of this model requires a tensor with infinitely many indices. This fits quite naturally into our framework.

We rewrite the partition function as a tensor network using the same method as for the Ising model in [14],[3, App. A.1], i.e. rotating the square lattice by 45 degrees. Specifically, we encode the interaction among $\theta_1, \theta_2, \theta_3, \theta_4$ at the vertices of a single plaquette by the function

$$A_{\theta_1\theta_2\theta_3\theta_4} = e^{\beta[\cos(\theta_1-\theta_2)+\cos(\theta_2-\theta_3)+\cos(\theta_3-\theta_4)+\cos(\theta_4-\theta_1)]}. \tag{3.15}$$

We view this function as a tensor on $V \times V \times V \times V$ where $V = L^2(S^1)$ with the measure $d\theta/(2\pi)$. The partition function Eq. (3.14) can then be obtained by placing copies of A in every other plaquette in a checkerboard pattern and contracting neighboring tensors where contraction means integrating the product over $\theta \in S^1$. The tensor A is Hilbert-Schmidt, the norm being given by

$$\|A\| = \left(\int \frac{d\theta_1}{2\pi} \frac{d\theta_2}{2\pi} \frac{d\theta_3}{2\pi} \frac{d\theta_4}{2\pi} |A_{\theta_1\theta_2\theta_3\theta_4}|^2 \right)^{1/2} < \infty\,. \tag{3.16}$$

To connect with the theory developed in this paper, we introduce an orthonormal basis in V and express A in this basis. We use the orthonormal basis $e_n = e^{in\theta}$, $n \in \mathbb{Z}$. Thus our Hilbert space basis is numbered by $\mathbb{Z}$, not by $\mathbb{Z}_{\geqslant 0}$ as in the previous sections. Hopefully this will not lead to a confusion. The index 0 will still play the same special role as in the previous sections. In this basis, our tensor is expressed as

$$A_{n_1n_2n_3n_4} = \int \frac{d\theta_1}{2\pi} \frac{d\theta_2}{2\pi} \frac{d\theta_3}{2\pi} \frac{d\theta_4}{2\pi} A_{\theta_1\theta_2\theta_3\theta_4}\, e^{-i(n_1\theta_1+n_2\theta_2+n_3\theta_3+n_4\theta_4)}. \tag{3.17}$$

To evaluate this, we use the Fourier series [16, (10.35.2)]:

$$e^{\beta\cos\theta} = \sum_{n \in \mathbb{Z}} I_n(\beta) e^{in\theta}, \tag{3.18}$$

where $I_n(\beta)$ is the modified Bessel function of the first kind. Substituting (3.18) into (3.15), we obtain after a straightforward computation:

$$A_{n_1 n_2 n_3 n_4} = \sum_{k\in\mathbb{Z}} I_k(\beta) I_{k+n_1}(\beta) I_{k+n_1+n_2}(\beta) I_{k+n_1+n_2+n_3}(\beta)\, \delta_{n_1+n_2+n_3+n_4}. \tag{3.19}$$

Note that the XY model is invariant under the $U(1)$ global symmetry, under which the state e_n has charge n. The $U(1)$ symmetry is visible in A through the charge conservation condition $n_1 + n_2 + n_3 + n_4 = 0$ for the nonzero tensor elements.

Let us note that Eq. (3.19) is not the only way to represent the XY model as a tensor network. Refs. [17, 18] provide an alternative method, which does not rotate the lattice by 45 degrees. It gives rise to a simpler-looking tensor

$$A^{\text{there}}_{n_1 n_2 n_3 n_4} = \sqrt{I_{n_1}(\beta) I_{n_2}(\beta) I_{n_3}(\beta) I_{n_4}(\beta)}\, \delta_{n_1+n_2+n_3+n_4}\,. \tag{3.20}$$

Both (3.19) and (3.20) are valid tensors representing the XY model partition function. They can be both used to get a bound on the high-T phase of the model. Below we work with Eq. (3.19), for two reasons. First, it was obtained by the same method we used of the Ising model, and it will be interesting to compare the result to the Ising model. Second, it happens to give a better bound than (3.20).[24]

We write

$$A = A_{0000}(A_* + b(\beta)), \tag{3.21}$$

where $b(\beta) \in \mathbb{H}_0$. We define $\mathbf{o}$ and $\mathbf{x}$ as follows:

$$\mathbf{o} = \operatorname{span}(e_0), \qquad \mathbf{x} = \operatorname{span}(\{e_n : n \neq 0\}). \tag{3.22}$$

We use these sectors on both vertical and horizontal legs of $b(\beta)$. As a first step, we need to compute a hat-tensor $\widehat{b}(\beta)$ for $b(\beta)$. $I_n(\beta)$ decreases exponentially with $|n|$. So the series (3.19) defining $A_{n_1 n_2 n_3 n_4}$ converges rapidly, and moreover the resulting tensor coefficients decrease exponentially for large n_1, n_2, n_3, n_4. For a purely numerical study, one could evaluate sufficiently many of these tensor coefficients, and thus get an approximate $\widehat{b}(\beta)$. However since here we are aiming for a rigorous result, we need a rigorously valid $\widehat{b}(\beta)$, which needs a bit of work to estimate tails. We describe our procedure in Appendix B.

Note that all states in the $\mathbf{x}$ sector carry non-zero $U(1)$ charge. So the hat tensor $\widehat{b}(\beta)$ satisfies the selection rule $\widehat{b}_{\mathbf{ooox}} = 0$ (and rotations).

We next do the preliminary RG step expressed by the same diagram (3.9) as for the Ising model analysis. This defines the tensor $A^{(0)} = A_* + b^{(0)}_{\text{XY}}(\beta)$ where $b^{(0)}_{\text{XY}}(\beta) \in \mathbb{H}_0$. The isometry on the horizontal legs acts as follows:

$$\mathbf{o}\otimes\mathbf{o} \to \mathbf{o}, \quad \mathbf{x}\otimes\mathbf{o} \to \mathbf{u}, \quad \mathbf{o}\otimes\mathbf{x} \to \mathbf{d}, \quad \mathbf{x}\otimes\mathbf{x} \to \mathbf{r}. \tag{3.23}$$

Thus, the horizontal Hilbert space of $A^{(0)}$ decomposes into 4 sectors $\mathbf{o}\otimes\mathbf{u}\otimes\mathbf{d}\otimes\mathbf{r}$. We obtain the hat-tensor $\widehat{b}^{(0)}_{\text{XY}}(\beta)$ from $\widehat{b}(\beta)$. Note that $(\widehat{b}^{(0)}_{\text{XY}})_{\mathbf{oooo}} = 0$ because $\widehat{b}_{\mathbf{ooox}} = 0$.

Remark 3.1. After the preliminary RG step, all states in sectors $\mathbf{x}, \mathbf{u}, \mathbf{d}$ (but not in $\mathbf{r}$) carry nonzero $U(1)$ charge. So $\widehat{b}^{(0)}_{\text{XY}}(\beta)$ like $\widehat{b}(\beta)$ satisfies a selection rule: elements with three indices $\mathbf{o}$ and the fourth index $\mathbf{x}$, $\mathbf{u}$ or $\mathbf{d}$ vanish. After we start acting with the 2x1 map, some states of zero charge will enter also into $\mathbf{x}, \mathbf{u}, \mathbf{d}$, and this selection rule will no longer apply. This is visible in Fig. 11 below. See also Section 2.11 and footnote 23.

[24] This was the result of our computations. A possible intuitive explanation is as follows. Because of the 45 degree lattice rotation, the tensor network built out of ℓ^2 A's corresponds to the XY partition function of $2\ell^2$ spins (see footnote 21 for the Ising). So passing from the spins to the tensors we have already performed a sort of "RG step" of scale factor $\sqrt{2}$. This is expected to bring us a bit closer to the high-T fixed point. Therefore, tensor RG iterations should have an easier time converging for A than for A^{there} for which there is no such effect.

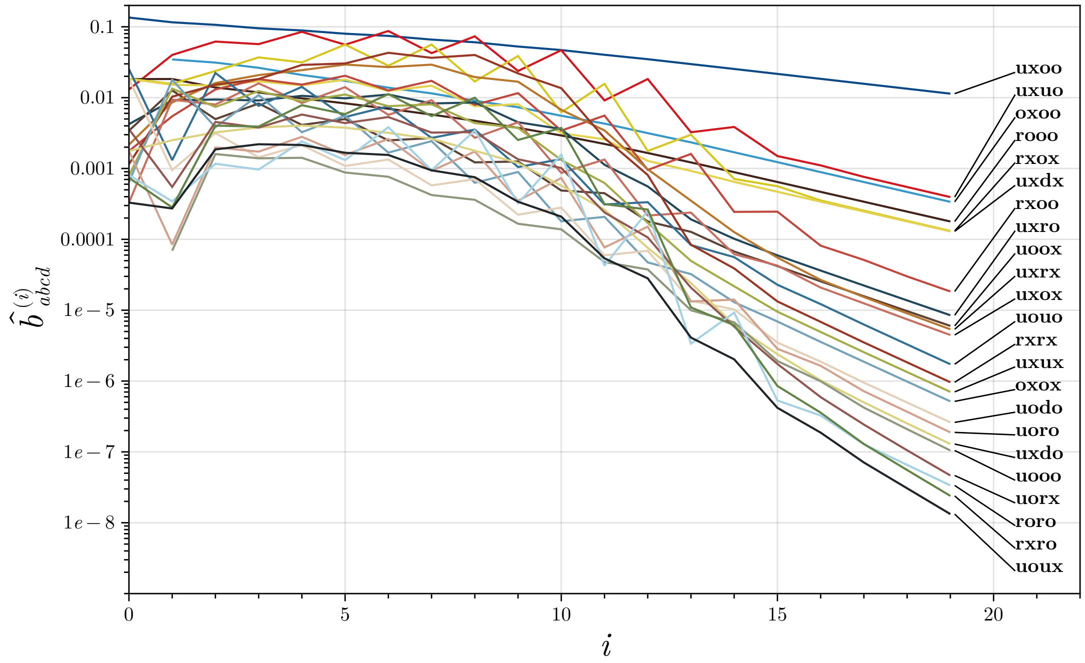


Figure 11: RG flow of the $\widehat{b}$ tensor for XY model initiated at $\beta = 0.19$. The 2x1 map parameters are: $w_{\mathbf{o}} = 2.4$ and $w_{\mathbf{x}} = 2.3$. Some of the trajectories appear only starting from the second step, as the initial tensor has these sectors identically zero (see Remark 3.1).

We then proceed like in the Ising model analysis, replacing $\widehat{b}^{(0)}_{\text{Ising}}(\bar{\beta})$ with $\widehat{b}^{(0)}_{\text{XY}}(\bar{\beta})$. We tweak the values of the reweighting parameters a bit to improve $\bar{\beta}$. We thus obtain:

Proposition 3.5. *(a) The set of tensors $\mathbb{O}(\widehat{b}^{(0)}_{\text{XY}}(\bar{\beta}))$, $\bar{\beta} = 0.19$, is a basin of stability for the 2x1 RG map with the reweighting parameters $w_{\mathbf{x}} = 2.3$, $w_{\mathbf{o}} = 2.4$.*

(b) The free energy is analytic on a neighborhood of this basin of stability.

(c) $b^{(0)}_{\text{XY}}(\beta) \in \mathbb{O}(\widehat{b}^{(0)}_{\text{XY}}(\bar{\beta}))$ for all $0 \leqslant \beta \leqslant \bar{\beta} - \delta$, $\delta = 5 \times 10^{-5}$.

Proof. For (a),(b) we proceed as in Proposition 3.4(a),(b) for the Ising model. Fig. 11 shows the corresponding master function iterates. The hypothesis of Lemma 2.3 is satisfied for $i_0 = 15$ with $\lambda = 0.86$ [8].

Part (c) is different from Ising. There, it was easy to see analytically that $\widehat{b}(\beta)$, and hence $\widehat{b}^{(0)}(\beta)$, were monotonic in β, so Proposition 3.4(c) was trivial with $\delta = 0$. For XY, we don't have an easy proof of this monotonicity, due to more complicated expressions of tensor elements. (Numerically we do see monotonicity.) So we prove (c) by a computer-assisted argument [8]. We split the interval $[0, \bar{\beta} - \delta]$ into finitely many intervals $I_i = [\beta_i, \beta_{i+1}]$. For each interval I_i we compute a tensor $\widehat{b}^{(0)}_i$ which is a hat tensor for all $b^{(0)}(\beta)$, $\beta \in I_i$. This is done by running the procedure described in App. B in interval arithmetic.[25] We then verify that $\widehat{b}^{(0)}_i \leqslant \widehat{b}^{(0)}(\bar{\beta})$ holds for all i. □

[25] To compute elements of the hat tensor $\widehat{b}(\beta)$ we need to evaluate the norm of tensor elements in each sector, and normalize dividing by A_{0000}. Since all tensor elements are monotonic, Eq. (3.19), we obtain a valid hat tensor for all $\beta \in I_i$ evaluating the norm at $\beta = \beta_{i+1}$ and dividing by A_{0000} at $\beta = \beta_i$. Our interval arithmetic code implements this logic automatically.

So we have proven that the XY model for $\beta \leqslant 0.18995$ lies in the high-T phase. The critical temperature for the XY model, where it transitions to a phase with algebraically decaying correlations (the BKT transition), is not exactly known. Various numerical techniques indicate that $\beta_c^{\mathrm{XY}} \approx 1.12$ (see e.g. [18, Table 1] or [19, Table I]), so our method captures approximately 17% of the high-T phase in this case. Note that the XY model is more disordered than the Ising model at the same β, in the sense that the spins have more room to fluctuate, and so it's natural that $\beta_c^{\mathrm{XY}} > \beta_c^{\mathrm{Ising}}$. It's reassuring that our method is also able to see this, as it gives a larger $\bar{\beta}$ for the XY than for the Ising. (There is also a rigorous bound $\beta_c^{\mathrm{XY}} \geqslant 2\beta_c^{\mathrm{Ising}}$ [20].)

4 Conclusions

We have introduced a new tensor RG map, but most importantly a new framework to do controlled renormalization group using tensor networks. The main characteristic of our RG map is that it comes equipped with a bounding box and a computable "master function" which shows how this bounding box varies from one RG step to the next. The RG flow happens in an infinite-dimensional space of tensors, but thanks to the bounding box we can control the RG flow via a finite calculation. Finding such a computer-assisted RG framework for lattice models has always been a dream, but it is perhaps for the first time that it is realized. This should allow myriad applications, once some generalizations described below are implemented.

In this paper we considered the simplest concrete application of this framework: proving that a lattice model is in the high-T phase. Of course there exist other means to do this. Perhaps the most standard tool is the cluster expansion, a.k.a. the polymer expansion [12]. Other methods include [21–23] and [24]. We are certain that our method, once it is further developed, has the potential to achieve larger domain of applicability than previous techniques. For now we were content to show that in its simplest form presented here it goes somewhat further than simple applications of the cluster expansion.

Let us discuss the possible ways to improve the domain of applicability of our approach. There are a few improvements that should be relatively straightforward to implement. When there was a contraction of a left tensor and a right tensor, we introduced reweighting factors $w_{\mathbf{x}}$ and $w_{\mathbf{o}}$ using the trivial identity $w\frac{1}{w} = 1$. A generalization is to introduce reweighting *matrices* $w = w_{AB}$ by inserting in the contraction the identity matrix in the form ww^{-1} or $w^{-1}w$. Indices A, B of these matrices will be of the form $A, B = abc$ where a, b, c are the sectors cut through by the red lines in Fig. 6 and Eq. (2.64). One can then optimize over invertible matrices w.

Our disentangler was chosen to cancel a certain class of diagrams to first order. Some such cancellation is essential to make the linearized map be a contraction, but the disentangler also produces terms that are not involved in this cancellation which can degrade our bounds. Using a less aggressive disentangler that does not fully cancel the selected diagrams could result in an improved domain of applicability.

We kept the parameters $w_{\mathbf{x}}$ and $w_{\mathbf{o}}$ constant along the RG trajectory. One can hope to improve the domain of applicability by allowing these parameters to vary for some finite number of initial steps of the RG flow. More generally, one can try varying other aspects of the RG map, e.g., the disentangler, for a finite number of RG steps.

Let us discuss now a more ambitious idea which should lead to a dramatic improvement of our method. In following our tensor along the RG trajectory, we wrote the tensor as $A_* + b$, and the only information we kept track of for b was the hat tensor $\widehat{b}$. The approach is to express the tensor in the form $A = A_* + b_f + b$ where b_f is a finite dimensional tensor and b is infinite dimensional. We express the image of A under the RG map in the form $A_* + b'_f + b'$ where b'_f is computed explicitly using a finite dimensional map that approximates the RG map. The infinite dimensional tensor b' is then determined by requiring that $A_* + b'_f + b'$ is the image of $A_* + b_f + b$. The tensor b is controlled by a hat tensor. The master function for that hat tensor is now a function of b_f, b. Since b contributes to the renormalization of b_f, tensor b'_f should be controlled by a set of intervals for its elements. After a finite number of iterations one can hope that b_f will be small enough that we can switch to the map studied in this paper.

Our long range goal is to use ideas from this paper to prove the existence of a critical fixed point tensor for some tensor RG map and control the map in a neighborhood of this fixed point. Here is a sketch of how such a proof might proceed. We express the tensor in the form $A = A_0 + b_f + b$ where A_0 and b_f are finite dimensional tensors and b is infinite dimensional. A_0 is an approximation to the critical fixed point. As above, the tensor b_f is computed explicitly using a finite dimensional map that approximates the RG map, and b' is determined by requiring that the RG map takes $A_0 + b_f + b$ to $A_0 + b'_f + b'$. (A_0 is kept fixed.) The tensor b is controlled by a hat tensor. Our experience with existing numerical RG maps make us optimistic that one can get an approximate fixed point A_0 that is easily within 0.02 of the exact fixed point. We would then prove that there exist b_f, b such that $A_0 + b_f + b$ is an exact fixed point. Most existing tensor network RG maps are not suitable for this program, in particular most use SVD which is not compatible with our HS norms. But we are hopeful that we can build a good RG map by modifying our 2x1 map.

Acknowledgments

SR and NE are supported in part by the Simons Foundation grant 733758 (Simons Bootstrap Collaboration). SR thanks the International Solvay Institute for Physics (Brussels) for hospitality while this work was being finalized. SR thanks the theoretical physicists of Belgium (as well as the architects, the painters, and the brewers) for the stimulating atmosphere. NE thanks Christophe Garban for the lecture series on BKT phase transitions that inspired the study of the XY model example for this paper.

A Accompanying code

All our main results were obtained with computer assistance. We provide the `Julia` [25, 26] code used in our analysis in the notebook `HT.ipynb`, which is available in our GitHub repository [8]. The main part of the code, defining the master function, comprises only about 300 lines of code distributed across notebook cells 3–5. We have thoroughly documented and commented the code, and we hope this will facilitate its understanding.

Readers who wish to check whether their favorite lattice model lies in the high-temperature phase can follow the instructions in the `README` file. A proof of such result requires only a few lines of code (see, for example, the first code cell in Section "General tensors," which proves Theorem 3.1).

Several packages made our results possible. Eq. (2.99) relies on `TensorOperations.jl` [27], `Symbolics.jl` [28], and `TaylorSeries.jl` [29]. The proofs presented in Section 3 depend on `TensorOperations.jl` [27], `ArbNumerics.jl` [30], `SpecialFunctions.jl` [31], and `HCubature.jl` [32, 33]. We use `CairoMakie.jl` [34] for plotting.

B $\widehat{b}(\beta)$ for the XY model

In this appendix we explain how we get a rigorous hat-tensor $\widehat{b}(\beta)$ for the XY tensor $b(\beta)$ defined by Eq. (3.21). We first present a series of lemmas containing estimates for the series defining $A_{n_1 n_2 n_3 n_4}$, and for the tails of the HS norm of A. We will then explain how we get $\widehat{b}(\beta)$ using those lemmas.

Recall some useful facts about the Bessel functions. We have $I_n(\beta) = I_{|n|}(\beta)$. The standard power series representation of $I_n(\beta)$ implies that $I_n(\beta) > 0$ for $\beta > 0$. For $n \geqslant 0$, it also implies a bound:

$$I_{n+1}(\beta) \leqslant \frac{\beta/2}{n+1} I_n(\beta)\,. \tag{B.1}$$

Using this several times, we get, for $n \geqslant 0$,

$$I_n(\beta) \leqslant \frac{(\beta/2)^n}{n!} I_0(\beta)\,. \tag{B.2}$$

We will sometimes write I_n for $I_n(\beta)$. In our estimates we will assume $\beta < 2$.

Lemma B.1. *Suppose $k_- \leqslant 0$, $k_+ \geqslant 0$ are such that the four integers $k, k+n_1, k+n_1+n_2, k+n_1+n_2+n_3$ are all $\geqslant 0$ for $k = k_+$ and all $\leqslant 0$ for $k = k_-$. Denote $W_k = I_k I_{k+n_1} I_{k+n_1+n_2} I_{k+n_1+n_2+n_3}$. Then, for $\beta < 2$,*

$$\sum_{k_-}^{k_+} W_k \leqslant A_{n_1 n_2 n_3 n_4} \leqslant \sum_{k_-}^{k_+} W_k + \frac{(\beta/2)^4}{1-(\beta/2)^4}(W_{k_+} + W_{k_-}). \tag{B.3}$$

Proof. The l.h.s. is obvious since all terms in the sum (3.19) defining $A_{n_1 n_2 n_3 n_4}$ are positive. For the r.h.s., terms for $k > k_+$ can be estimated using (B.1) by $(\beta/2)^{4(k-k_+)}$ times W_{k_+}. Analogously for terms with $k < k_-$. □

Lemma B.2. *Denote $n = \max(|n_1|, |n_2|, |n_3|, |n_4|)$. Then, for $\beta < 2$,*

$$A_{n_1 n_2 n_3 n_4} \leqslant \frac{(I_0(\beta))^4}{1-(\beta/2)^4} \frac{4n+1}{\Gamma(n/2+1)^2} (\beta/2)^n\,. \tag{B.4}$$

Proof. We pick $k_\pm = \pm 2n$. These $k_\pm$ satisfy the requirements of the previous lemma (note that $n_1 + n_2 + n_3 = -n_4$ by charge conservation). So we can use the upper bound from that lemma. Consider any W_k for $|k| \leqslant 2n$. Suppose for definiteness that n_1 is one of the maximal indices, i.e. $n = |n_1|$ (other cases are similar). Other indices n_2, n_3, n_4 may be anywhere in the range $[-n, n]$. Suppose also for definiteness that $n_1 \geqslant 0$. Then we estimate

$$W_k \leqslant I_k I_{k+n} I_0^2 \leqslant \frac{(\beta/2)^{|k|+|k+n|}}{|k|!|k+n|!} I_0^4, \tag{B.5}$$

where in the second inequality we used (B.2). The maximum of the r.h.s. is realized for $k = -n/2$. Estimating all terms in $k \in [-2n, 2n]$ by the maximum, and adding the last term in (B.3), we get the stated result. □

Lemma B.3. *For a positive integer n, define the set T_n of triples of integers (x, y, z) which are all not larger than n in absolute value and whose sum equals n:*

$$T_n = \{(x, y, z) \in \mathbb{Z}^3 : x, y, z \in [-n, n], x + y + z = n\} \tag{B.6}$$

Then $|T_n| = 2n^2 + 3n + 1$.

This is an elementary counting exercise. We omit the proof.

Lemma B.4. *Suppose $N \geqslant 3$. Then, for $\beta < 2$,*

$$\sum_{\max(|n_1|,|n_2|,|n_3|,|n_4|)\geqslant N} |A_{n_1 n_2 n_3 n_4}|^2 \leqslant 4K_N \left[\frac{(I_0(\beta))^4}{1-(\beta/2)^4}\right]^2 \frac{(\beta/2)^{2N}}{1-(\beta/2)^2}, \tag{B.7}$$

where $K_n = 2|T_n| \left(\frac{4n+1}{\Gamma(n/2+1)^2}\right)^2$.

Proof. We split the sum into 4 groups corresponding to which $|n_i|$ is largest (the groups overlap because several $|n_i|$ may be equal). Consider for definiteness the group where $|n_4|$ is largest. Many terms in the group are zero because they don't satisfy charge conservation. By Lemma B.3, the number of nonzero terms with a fixed $|n_4| = n$ is $2|T_n|$ (the factor 2 accounts for two signs of n_4). Bounding each nonzero term by Lemma B.2, the sum of terms in the group is bounded by

$$\left[\frac{(I_0(\beta))^4}{1-(\beta/2)^4}\right]^2 \sum_{n\geqslant N} K_n(\beta/2)^{2n} \leqslant K_N \left[\frac{(I_0(\beta))^4}{1-(\beta/2)^4}\right]^2 \frac{(\beta/2)^{2N}}{1-(\beta/2)^2}, \tag{B.8}$$

where we used that K_n is monotonically decreasing for $n \geqslant 3$ as it is easy to check. We then multiply this estimate by 4 to account for 4 groups. □

The nonzero components of the minimal hat-tensor $\widehat{b} = \widehat{b}(\beta)$ are given by

$$\widehat{b}_{abcd} = (A_{0000})^{-1}\|A_{abcd}\|, \quad abcd = \mathbf{ooxx}, \mathbf{oxox}, \mathbf{oxxx}, \mathbf{xxxx}\,, \tag{B.9}$$

and rotations thereof (the tensor $\widehat{b}$ is rotation-invariant). To get a hat-tensor we have to provide upper bounds for all quantities in the r.h.s. These bounds are obtained by Lemmas B.1,B.4. Moreover we can get bounds close to the exact values by playing with the parameters $k_\pm, N$ in those lemmas. To make this clear let $\epsilon > 0$ be an accuracy parameter. In practice maybe $\epsilon = 10^{-16}$ the machine precision. The bounds described below all go to the exact values as $\epsilon \to 0$.

A lower bound on A_{0000} (which gives an upper bound on its inverse) is given by the l.h.s. of (B.3) taking $k_\pm$ so large that the r.h.s. and the l.h.s. differ by less than ϵ.

Let A_{abcd} be any of the A sectors in the r.h.s. of (B.9). We choose N so large that the r.h.s. of (B.4) is smaller than ϵ. We then bound $\|A_{abcd}\|$ by

$$\|A_{abcd}\| \leqslant \left(\sum_{\substack{n_1\in a, n_2\in b, n_3\in c, n_4\in d\\ |n_1|,|n_2|,|n_3|,|n_4|<N}} |A_{n_1n_2n_3n_4}|^2 + \epsilon\right)^{1/2}. \tag{B.10}$$

So we are reduced to providing upper bounds for each $A_{n_1n_2n_3n_4}$ in the r.h.s. of (B.10). Since N is fixed this is a finite number of tensor elements to consider. We bound them by the r.h.s. of (B.3) taking $k_\pm$ as large as that lemma requires so that it applies, and so that the two sides of (B.3) differ by less than ϵ. We use interval arithmetic. In particular the Bessel functions $I_n(\beta)$ are evaluated in interval arithmetic. This is how we get a rigorous hat-tensor $\widehat{b}(\beta)$.

References

[1] M. Levin and C. P. Nave, "Tensor renormalization group approach to 2D classical lattice models," *Phys. Rev. Lett.* **99** no. 12, (2007) 120601, arXiv:cond-mat/0611687.

[2] G. Evenbly, "Algorithms for tensor network renormalization," arXiv:1509.07484 [cond-mat.str-el].

[3] T. Kennedy and S. Rychkov, "Tensor RG Approach to High-Temperature Fixed Point," *J. Statist. Phys.* **187** no. 3, (2022) 33, arXiv:2107.11464 [math-ph].

[4] T. Kennedy and S. Rychkov, "Tensor Renormalization Group at Low Temperatures: Discontinuity Fixed Point," *Annales Henri Poincare* **25** no. 1, (2024) 773–841, arXiv:2210.06669 [math-ph].

[5] N. Ebel, "3D Tensor Renormalisation Group at High Temperatures," *Annales Henri Poincare* (July, 2024) , arXiv:2401.04229 [cond-mat.stat-mech].

[6] N. Ebel, T. Kennedy, and S. Rychkov, "Rotations, Negative Eigenvalues, and Newton Method in Tensor Network Renormalization Group," arXiv:2408.10312 [cond-mat.stat-mech].

[7] N. Ebel, T. Kennedy, and S. Rychkov, "Transfer Matrix and Lattice Dilatation Operator for High-Quality Fixed Points in Tensor Network Renormalization Group," arXiv:2409.13012 [cond-mat.stat-mech].

[8] https://github.com/ebelnikola/HT.

[9] G. Evenbly and G. Vidal, "Tensor Network Renormalization," *Phys. Rev. Lett.* **115** no. 18, (2015) 180405, arXiv:1412.0732 [cond-mat.str-el].

[10] G. Vidal, "Entanglement Renormalization," *Phys. Rev. Lett.* **99** no. 22, (2007) 220405, arXiv:cond-mat/0512165.

[11] D. Ruelle, *Statistical Mechanics: Rigorous Results*. World Scientific, 1999.

[12] S. Friedli and Y. Velenik, *Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction*. Cambridge University Press, 2017.

[13] R. Kotecký and D. Preiss, "Cluster expansion for abstract polymer models," *Commun. Math. Phys.* no. 103, (1986) 491–498.

[14] Z.-C. Gu and X.-G. Wen, "Tensor-entanglement-filtering renormalization approach and symmetry-protected topological order," *Phys. Rev. B* **80** no. 15, (2009) 155131, arXiv:0903.1069 [cond-mat.str-el].

[15] B. Simon, *The Statistical Mechanics of Lattice Gases, Volume I.* 2014.

[16] "*NIST Digital Library of Mathematical Functions*." https://dlmf.nist.gov/, release 1.2.4 of 2025-03-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[17] J. F. Yu, Z. Y. Xie, Y. Meurice, Y. Liu, A. Denbleyker, H. Zou, M. P. Qin, J. Chen, and T. Xiang, "Tensor renormalization group study of classical XY model on the square lattice," *Phys. Rev. E* **89** (Jan, 2014) 013308, arXiv:1309.4963 [cond-mat.stat-mech].

[18] R. G. Jha, "Critical analysis of two-dimensional classical XY model," *Journal of Statistical Mechanics: Theory and Experiment* **2020** no. 8, (Aug, 2020) 083203, arXiv:2004.06314 [hep-lat].

[19] A. Ueda and M. Oshikawa, "Resolving the Berezinskii-Kosterlitz-Thouless transition in the two-dimensional XY model with tensor-network-based level spectroscopy," *Phys. Rev. B* **104** no. 16, (Oct., 2021) 165132, arXiv:2105.11460 [cond-mat.stat-mech].

[20] M. Aizenman and B. Simon, "A comparison of plane rotor and Ising models," *Physics Letters A* **76** no. 3, (1980) 281–282.

[21] R. L. Dobrushin and S. B. Shlosman, "Completely Analytical Gibbs Fields," in *Statistical Physics and Dynamical Systems: Rigorous Results*, Y. M. Sinai, ed., pp. 371–403. Birkhäuser Boston, Boston, MA, 1985.

[22] R. L. Dobrushin, J. Kolafa, and S. B. Shlosman, "Phase Diagram of the Two-Dimensional Ising Antiferromagnet (Computer-Assisted Proof)," *Communications in Mathematical Physics* **102** (1985) 89–103.

[23] R. L. Dobrushin and S. B. Shlosman, "Completely Analytical Interactions: Constructive Description," *Journal of Statistical Physics* **46** no. 5–6, (1987) 983–1014.

[24] T. Kennedy, "A Fixed-Point Equation for the High-Temperature Phase of Discrete Lattice Spin Systems," *Journal of Statistical Physics* **59** no. 1–2, (1990) 1–29.

[25] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A fresh approach to numerical computing," *SIAM Review* **59** no. 1, (2017) 65–98.

[26] "Julia." https://julialang.org/.

[27] L. Devos, M. Van Damme, J. Haegeman, and contributors, "TensorOperations.jl," Oct, 2023. https://github.com/Jutho/TensorOperations.jl.

[28] S. Gowda, Y. Ma, A. Cheli, M. Gwozdz, V. B. Shah, A. Edelman, and C. Rackauckas, “High-performance symbolic-numerics via multiple dispatch,” arXiv:2105.03949 [cs.CL].

[29] L. Benet and D. P. Sanders, “TensorSeries.jl,” Apr, 2025. https://github.com/JuliaDiff/TaylorSeries.jl.

[30] J. Sarnoff and contributors, “ArbNumerics.jl.” https://github.com/JeffreySarnoff/ArbNumerics.jl.

[31] “SpecialFunctions.jl.” https://github.com/JuliaMath/SpecialFunctions.jl.

[32] S. G. Johnson, “HCubature.jl.” https://github.com/JuliaMath/HCubature.jl, 2017.

[33] A. C. Genz and A. A. Malik, “Remarks on algorithm 006: An adaptive algorithm for numerical integration over an n-dimensional rectangular region,” *Journal of Computational and Applied Mathematics* **6** (1980) 295–302.

[34] S. Danisch and J. Krumbiegel, “Makie.jl: Flexible high-performance data visualization for Julia,” *Journal of Open Source Software* **6** no. 65, (2021) 3349.