Journal of Machine Learning Research 23 (2022) 1-36 Submitted 2/21; Revised 6/22; Published 8/22

Mappings for Marginal Probabilities with Applications to Models in Statistical Physics

Mehdi Molkaraie MEHDI.MOLKARAIE@ALUMNI.ETHZ.CH
Department of Statistical Sciences
University of Toronto
Toronto ON M5G 1Z5
Canada

Editor: Mohammad Emtiyaz Khan

Abstract

We present local mappings that relate the marginal probabilities of a global probability mass function represented by its primal normal factor graph to the corresponding marginal probabilities in its dual normal factor graph. The mapping is based on the Fourier transform of the local factors of the models. Details of the mapping are provided for the Ising model, where it is proved that the local extrema of the fixed points are attained at the phase transition of the two-dimensional nearest-neighbor Ising model. The results are further extended to the Potts model, to the clock model, and to Gaussian Markov random fields. By employing the mapping, we can transform simultaneously all the estimated marginal probabilities from the dual domain to the primal domain (and vice versa), which is advantageous if estimating the marginals can be carried out more efficiently in the dual domain. An example of particular significance is the ferromagnetic Ising model in a positive external magnetic field. For this model, there exists a rapidly mixing Markov chain (called the subgraphs–world process) to generate configurations in the dual normal factor graph of the model. Our numerical experiments illustrate that the proposed procedure can provide more accurate estimates of marginal probabilities of a global probability mass function in various settings.

Keywords: Normal Factor Graph, Factor Graph Duality, Ising Model, Potts Model, Gaussian Markov Random Fields, Subgraphs–World Process, Phase Transition, Marginal Probability, Fourier Transform, Monte Carlo Methods.

1. Introduction

One of the main objectives in any probabilistic inference problem is computing local marginal probabilities of a global multivariate probability mass function (PMF). In general, such a computation requires a summation with an exponential number of terms, which makes its exact computation intractable (Wainwright and Jordan, 2008, Chapter 2). See also (Dagum and Luby, 1993) for the hardness results on approximating conditional probabilities in the context of belief networks.

In this paper, our approach for estimating the local marginal probabilities of a global PMF hinges on the notions of normal realization, in which there is an edge for every variable (Forney, 2001), normal factor graph (NFG), and the NFG duality theorem (Al-

Bashabsheh and Mao, 2011). A succinct discussion on factor graphs and NFGs is provided in Appendix A.

In our analysis, we mainly focus on well-known models in Statistical Physics. Specifically, we consider binary models with symmetric pairwise nearest-neighbour interactions (e.g., the Ising model). We will also consider extensions of our results to non-binary models (e.g., the q-state Potts model and the clock model), and briefly discuss generalizations to continuous models (e.g., Gaussian Markov random fields). These models are widely used in machine learning (Murphy, 2012), image processing (Winkler, 1995; McGrory et al., 2009; Liu et al., 2022), and spatial statistics (Besag and Green, 1993). Furthermore, the well-known models in Statistical Physics can be used as the basis to analyze less conventional and more complex models. Connections between Statistical Physics and Machine Learning run deep. For a comprehensive review, see (Carleo et al., 2019) and references therein.

The NFG duality theorem states that the partition function of a primal NFG and the partition function of its dual NFG are equal up to some known scale factor. It has been demonstrated that, in the low-temperature regime, Monte Carlo methods for estimating the partition function converge faster in the dual NFG than in the primal NFG of the two-dimensional (2D) Ising model (Molkaraie and Loeliger, 2013) and of the 2D Potts model (Al-Bashabsheh and Mao, 2014; Molkaraie and Gómez, 2018). This work extends the previous results to marginal probabilities of a primal NFG and its dual.

A preliminary version of this work was presented at (Molkaraie, 2020), where the NFG duality theorem was employed to establish a connection between the marginal probabilities of a global PMF associated with a primal NFG and their corresponding marginals in its dual NFG via local mappings. In this paper, we give a more detailed analysis on the proposed mappings – specially for non-binary and for continuous models. New numerical experiments are also provided for continuous graphical models.

The mapping can be expressed in terms of the discrete Fourier transform (DFT) of the local factors of the graphical models. Furthermore, the mappings are independent of the size of the model, of the topology of the graphical model, and of any assumptions on the parameters of the model (e.g., the coupling parameters).

The proposed method is practically beneficial if estimating marginal probabilities can be carried out more efficiently in the dual domain than in the primal domain. Indeed, there is a rapidly mixing Markov chain (called the subgraphs-world process) to generate configurations in the dual NFG of ferromagnetic Ising models with arbitrary topology and in a positive external field. The marginal probabilities in the dual domain can thus be estimated from such configurations. For more details see (Jerrum and Sinclair, 1993).

In a related topic and in the context of Coding Theory, it has been demonstrated that, for a class of binary symmetric channels, the correlation between code bits of low-density parity-check (LDPC) codes and low-density generator-matrix (LDGM) codes decays exponentially fast in the high-noise and the low-noise regimes, respectively. For more details see (Kudekar and Macris, 2011) and references therein. LDGM codes can be viewed as the dual of LDPC codes, and vice versa (Richardson and Urbanke, 2008, Chapter 7).

As an example, let us consider a 2D homogeneous Ising model with periodic boundary conditions. In this model, all marginal probabilities can be expressed as the ratio of two partition functions. In the high-temperature regime, the ratio can be estimated efficiently in the primal NFG, and at low temperatures, the ratio can be estimated efficiently in the

dual NFG of the model (Molkaraie and Loeliger, 2013; Al-Bashabsheh and Mao, 2014; Molkaraie and Gómez, 2018). In more general settings (e.g., in non-homogeneous models), each marginal probability needs to be estimated *separately* as the ratio of two partition functions. However, the mappings of this paper allow a *simultaneous* transformation of estimated marginal probabilities from one domain to the other. In the dual domain, the marginal probabilities can be estimated via variational inference algorithms (e.g., the belief propagation (BP) and the tree expectation propagation (TEP) algorithms), via Markov chain Monte Carlo methods, or in the special case of the ferromagnetic Ising models via the subgraphs-world process.

The paper is organized as follows. In Sections 2 and 3 we describe our models in the primal domain and discuss their graphical representations in terms of NFGs. Section 4 discusses the dual NFG, the NFG duality theorem, and the high-temperature series expansion of the partition function of the Ising model. In Section 5, we derive mappings that relate the edge/vertex marginal probabilities in the primal NFG to the corresponding edge/vertex marginal probabilities in the dual NFG. For binary models, details of the mapping, the fixed points of the mapping, and its connections to local magnetization are discussed in Section 6. Generalizations of the mapping and its fixed points to non-binary models are discussed in Sections 7. Numerical experiments for 2D and fully-connected graphical models are reported in Section 8. Finally, section 9 briefly discusses extensions of the mappings to continuous (Gaussian) models. Appendix A is a compact tutorial on factor graphs, NFGs, and the NFG duality. In Appendix B, the mapping is employed as an alternative method to compute the marginal probabilities of the 1D Ising and Potts models.

2. The Primal Model

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a finite, simple, connected, and undirected graph with $|\mathcal{V}|$ vertices (sites) and $|\mathcal{E}|$ edges (bonds). Suppose random variables $\mathbf{X} = (X_v, v \in \mathcal{V})$ and $\mathbf{Y} = (Y_e, e \in \mathcal{E})$ are indexed by the elements of $\mathcal{V}$ and $\mathcal{E}$, respectively. We refer to a vertex by X_v (i.e., the random variable associated with v) or for brevity by its index v. Similarly, we may refer to the edge that connects X_k and X_ℓ by (k, ℓ), by Y_e, or simply by its index e.

Each variable takes values in a finite alphabet $\mathcal{A}$, where $\mathcal{A} = \mathbb{Z}/q\mathbb{Z}$, i.e., the ring of integers modulo q, for some fixed integer $q \geq 2$. However, we may view $\mathcal{A}$ as any finite abelian group with respect to addition. An assignment of values to $\mathbf{X}$ (to $\mathbf{Y}$) is denoted by $\mathbf{x}$ (by $\mathbf{y}$) and is called a *configuration*. Here, $\mathbf{x} \in \mathcal{A}^{|\mathcal{V}|}$ and $\mathbf{y} \in \mathcal{A}^{|\mathcal{E}|}$ are column vectors of length $|\mathcal{V}|$ and $|\mathcal{E}|$, respectively.

For a subset $\mathcal{I}$ of $\mathcal{V}$ we let $\mathbf{X}_\mathcal{I} = (X_i, i \in \mathcal{I})$. We further define two indicator functions: the equality indicator function, which is defined as follows

$$\delta_{=}(\mathbf{x}_\mathcal{I}) = \begin{cases} 1, & \text{if } x_1 = x_2 = \ldots = x_{|\mathcal{I}|} \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

and the zero-sum indicator function, which is defined as

$$\delta_{+}(\mathbf{x}_\mathcal{I}) = \begin{cases} 1, & \text{if } x_1 + x_2 + \ldots + x_{|\mathcal{I}|} = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

For $|\mathcal{I}| = 1$, both (1) and (2) are equivalent to the Kronecker delta function, denoted by $\delta(\cdot)$. Notice that $\delta_{=}(x_k, x_\ell)$ and $\delta_{+}(x_k, x_\ell)$ are both equal to $\delta(x_k - x_\ell)$ when $\mathcal{A} = \mathbb{Z}/2\mathbb{Z}$ (i.e., the binary field).

In this paper, we assume that variables have pairwise interactions. Two variables interact if their corresponding vertices are connected by an edge in $\mathcal{G}$. Following Forney (2018), we suppose $\boldsymbol{M}$ is an oriented incidence matrix of $\mathcal{G}$ with size $|\mathcal{E}| \times |\mathcal{V}|$. To construct $\boldsymbol{M}$, we first give each edge an arbitrary orientation. We then set the entry $\boldsymbol{M}_{e,v} = +1$ if Y_e leaves X_v, the entry $\boldsymbol{M}_{e,v} = -1$ if Y_e enters X_v, and the entry $\boldsymbol{M}_{e,v} = 0$ otherwise. Therefore, each row of $\boldsymbol{M}$ has exactly two nonzero entries, and the number of nonzero entries in the v-th column of $\boldsymbol{M}$ is equal to the degree of X_v. Since $\mathcal{G}$ is connected, the rank of $\boldsymbol{M}$ is $|\mathcal{V}| - 1$.

In this framework

$$\mathbf{y} \colon \mathcal{A}^{|\mathcal{V}|} \to \mathcal{A}^{|\mathcal{E}|}, \quad \mathbf{x} \mapsto \boldsymbol{M}\mathbf{x} \tag{3}$$

Each entry y_e of $\mathbf{y}(\mathbf{x}) = \boldsymbol{M}\mathbf{x}$ can thus be expressed as the difference between x_k and x_ℓ, where X_k and X_ℓ are connected by Y_e. In the sequel, we drop the argument $\mathbf{x}$, and use $\mathbf{y}$ instead of $\mathbf{y}(\mathbf{x})$ when there is no ambiguity.

In the primal domain, we assume that the probability of a configuration $\mathbf{x} \in \mathcal{A}^{|\mathcal{V}|}$ is given by the following PMF

$$\pi_{\mathrm{p}}(\mathbf{x}) = \frac{1}{Z_{\mathrm{p}}} \prod_{e \in \mathcal{E}} \psi_e(y_e) \prod_{v \in \mathcal{V}} \phi_v(x_v) \tag{4}$$

for some set of *edge-weighing factors* $\{\psi_e \colon \mathcal{A} \to \mathbb{R}_{\geq 0}, e \in \mathcal{E}\}$ and *vertex-weighing factors* $\{\phi_v \colon \mathcal{A} \to \mathbb{R}_{\geq 0}, v \in \mathcal{V}\}$. The normalization constant Z_{p} is the *partition function*, which can be computed as

$$Z_{\mathrm{p}} = \sum_{\mathbf{x} \in \mathcal{A}^{|\mathcal{V}|}} \prod_{e \in \mathcal{E}} \psi_e(y_e) \prod_{v \in \mathcal{V}} \phi_v(x_v), \tag{5}$$

where the sum runs over all the vectors in the configuration space $\mathcal{A}^{|\mathcal{V}|}$.

In (4) we have assumed that each edge-weighing factor $\psi_e(\cdot)$ is only a function of the edge configuration y_e. In the sequel, we show that many important models in Statistical Physics (e.g., the Ising model and the q-state Potts model) can be easily represented in this framework. In Section 9, we will show that a Gaussian Markov random field with the thin-membrane prior can also be expressed by (4).

2.1 Examples from Statistical Physics

In the Ising and Potts models variables (particles) are associated with the vertices of $\mathcal{G}$, and two particles interact if they are joined by and edge. A real coupling parameter is associated with each interacting pair of particles. The model is called *ferromagnetic* if coupling parameters are nonnegative. If coupling parameters are constant, the model is called *homogeneous*. Moreover, particles may be under the influence of an external magnetic field. The external field is usually assumed to be constant. However, in more general settings, the field may vary from site to site.

The energy of a configuration $\mathbf{x} \in \mathcal{A}^{|\mathcal{V}|}$ is given by the Hamiltonian $\mathcal{H}(\mathbf{x})$. The probability of a configuration $\mathbf{x}$ is given by the Gibbs-Boltzmann distribution $\pi \propto \mathrm{e}^{-\beta \mathcal{H}(\mathbf{x})}$, where β is proportional to the inverse temperature (Huang, 1987). We sometimes denote exp(1) by e to distinguish it from the variable e that we use to denote the edges of $\mathcal{G}$.

- **The Ising model.** In an Ising model $\mathcal{A} = \mathbb{Z}/2\mathbb{Z}$, and each particle may be in one of the two states (e.g., spin up or spin down). The Hamiltonian of the model is given by[1]

$$\mathcal{H}(\mathbf{x}) = -\sum_{(k,\ell)\in\mathcal{E}} J_{k,\ell}\left(2\delta_{=}(x_k, x_\ell) - 1\right) - \sum_{v\in\mathcal{V}} H_v\left(2\delta(x_v) - 1\right) \tag{6}$$

  Here $J_{k,\ell}$ and H_v denote the coupling parameter associated with the interacting pair (X_k, X_ℓ) and the external field at site v, respectively.

  Starting from the Gibbs-Boltzmann distribution, it is straightforward to verify that the model can be represented by the PMF (4), in which the edge-weighing factors are

$$\psi_e(y_e) = \begin{cases} \mathrm{e}^{\beta J_e}, & \text{if } y_e = 0 \\ \mathrm{e}^{-\beta J_e}, & \text{if } y_e = 1, \end{cases} \tag{7}$$

  where $\mathbf{y} = \boldsymbol{M}\mathbf{x}$ and J_e is the coupling parameter associated with $e = (k, \ell)$, and the vertex-weighing factors are given by

$$\phi_v(x_v) = \begin{cases} \mathrm{e}^{\beta H_v}, & \text{if } x_v = 0 \\ \mathrm{e}^{-\beta H_v}, & \text{if } x_v = 1 \end{cases} \tag{8}$$

- **The q-state Potts model.** In the Potts model each variable represents the q possible states of a particle, which takes values in $\mathcal{A} = \mathbb{Z}/q\mathbb{Z}$. The Hamiltonian of the model reads

$$\mathcal{H}(\mathbf{x}) = -\sum_{(k,\ell)\in\mathcal{E}} J_{k,\ell}\delta_{=}(x_k, x_\ell) - \sum_{v\in\mathcal{V}} H_v\delta(x_v) \tag{9}$$

  Here, following (Nishimori and Ortiz, 2015, Chapter 1), we have assumed that the external field affects the variable x_v only if $x_v = 0$.

  Again, from the Gibbs-Boltzmann distribution, it is easy to show that the model can be represented by (4) with

$$\psi_e(y_e) = \begin{cases} \mathrm{e}^{\beta J_e}, & \text{if } y_e = 0 \\ 1, & \text{otherwise,} \end{cases} \tag{10}$$

  and

$$\phi_v(x_v) = \begin{cases} \mathrm{e}^{\beta H_v}, & \text{if } x_v = 0 \\ 1, & \text{otherwise.} \end{cases} \tag{11}$$

- **The clock model.** In the absence of an external field, let

$$\mathcal{H}(\mathbf{x}) = -\sum_{(k,\ell)\in\mathcal{E}} J_{k,\ell}\cos(2\pi(x_k - x_\ell)/q), \tag{12}$$

  where $\mathcal{A} = \mathbb{Z}/q\mathbb{Z}$. This model is known as the clock model (or the vector Potts model), which reduces to the Ising model if $q = 2$ and to the three-state Potts model if $q = 3$. In the limit of large q, the model is equivalent to the XY model.

1. In the bipolar case (i.e., when $\mathcal{A} = \{-1, +1\}$), the Hamiltonian of the Ising model has the more familiar form $\mathcal{H}(\mathbf{x}) = -\sum_{(k,\ell)\in\mathcal{E}} J_{k,\ell}x_k x_\ell - \sum_{v\in\mathcal{V}} H_v x_v$. For more details, see (Baxter, 2007, Chapter 1).

The clock model can also be expressed by (4). E.g., the edge-weighing factors of a four-state clock model with Hamiltonian (12) are given by

$$\psi_e(y_e) = \begin{cases} \mathrm{e}^{\beta J_e}, & \text{if } y_e = 0 \\ \mathrm{e}^{-\beta J_e}, & \text{if } y_e = 2 \\ 1, & \text{otherwise.} \end{cases} \tag{13}$$

For more details, see (Yeomans, 1992; Baxter, 2007).

Next, we will discuss the graphical representation of (4) in terms of NFGs.

3. The Primal NFG

The factorization in (4) can be represented by an NFG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, in which vertices represent the factors and edges represent the variables. The edge that represents some variable y_e is connected to the vertex representing the factor $\psi_e(\cdot)$ if and only if y_e is an argument of $\psi_e(\cdot)$. If a variable is involved in more than two factors, it is replicated via an equality indicator factor. See Appendix A for more details.

The primal NFG of the 1D Potts model in an external field with periodic boundary conditions is depicted in Fig. 1, where the big unlabeled boxes (attached to Y_1, Y_2, and Y_3) represent (10), and the small unlabeled boxes (attached to X_1, X_2, and X_3) represent (11).

In Fig. 1, boxes labeled "=" are instances of equality indicator factors given by (1), which impose the constraint that all their incident variables are equal, e.g., the equality indicator factor $\delta_=(x_1, x_1', x_1'')$ is equal to one if $x_1 = x_1' = x_1''$, and is equal to zero otherwise. In analogy with the inversion bubble connected to the logic NOT, NAND, and NOR gate symbols (Horowitz and Hill, 1989), the symbol "$\circ$" is used to indicate a sign inversion. The boxes labeled "+" are instances of zero-sum indicator factors given by (2). They impose the constraint that all their incident variables sum to zero, e.g., the zero-sum indicator factor $\delta_+(y_1, x_1', x_2')$ is equal to one if $y_1 + x_1' + x_2' = 0$, and evaluates to zero otherwise. (Recall that all arithmetic manipulations are modulo q.)

In the primal domain, the NFG of a 1D Ising model with periodic boundary conditions is also shown in Fig. 1, where the big unlabeled boxes represent (7), and the small unlabeled boxes represent (8). In the case of the Ising model $\mathcal{A} = \mathbb{Z}/2\mathbb{Z}$ (i.e., the binary field), the "$\circ$" symbols are irrelevant and can be removed from the primal NFG. (Note that in the binary field $x_1 + x_1 = 0$, and therefore $-x_1 = x_1$.)

As discussed in Section 2, we observe that in Fig. 1 we can freely choose $\mathbf{X}$ and therefrom fully determine $\mathbf{Y}$. More generally, if we take $\mathcal{G}$ to be a d-dimensional model with pairwise interactions, we can compute each component Y_e of $\mathbf{Y}$ from the two components of $\mathbf{X}$ that are incident to the zero-sum indicator factor attached to Y_e.

4. The Dual NFG

We can obtain the dual of a primal NFG representing (4) by employing the following steps.

- Replace each variable, say X_k, by its corresponding dual variable $\tilde{X}_k$.
- Replace each factor $\psi_e(\cdot)$ by its 1D DFT $\tilde{\psi}_e(\cdot)$.

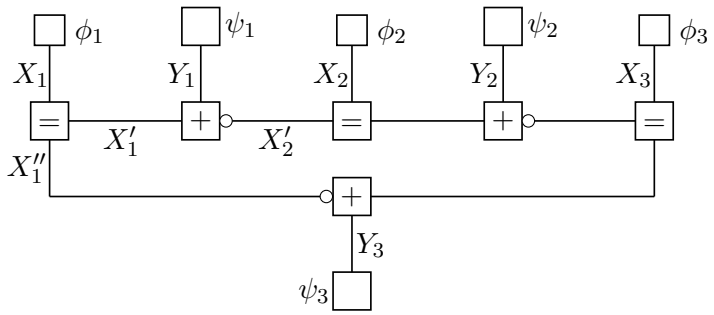

Figure 1: The primal NFG of the 1D Potts model in an external field with periodic boundary conditions, where the big unlabeled boxes represent (10), and the small unlabeled boxes represent (11). Boxes labeled "=" are instances of equality indicator factors as in (1), the boxes labeled "+" are instances of zero-sum indicator factors as in (2), and the symbol "∘" denotes a sign inversion.

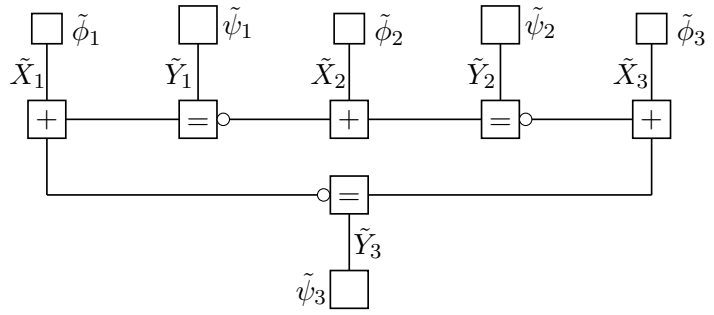

Figure 2: The dual NFG of Fig. 1, where the big unlabeled boxes represent (17), and the small unlabeled boxes represent (18). Boxes labeled "=" are instances of equality indicator factors as in (1), the boxes labeled "+" are instances of zero-sum indicator factors as in (2), and the symbol "∘" models a sign inversion.

- Replace each factor $\phi_v(\cdot)$ by its 1D DFT $\tilde{\phi}_v(\cdot)$.
- Replace equality indicator factors by zero-sum indicator factors, and vice-versa.

The above dualizations procedure retains the topology of the primal NFG. In general, dual variables take values in the dual of $\mathcal{A}$ (i.e., its character group). However, the dual group can be taken as $\mathcal{A} = \mathbb{Z}/q\mathbb{Z}$ in our framework. We will use the tilde symbol to denote variables in the dual domain.

Here $\tilde{\psi}_e(\cdot)$ the 1D DFT of $\psi_e(\cdot)$ can be computed as

$$\tilde{\psi}_e(\tilde{y}_e) = \sum_{y \in \mathcal{A}} \psi_e(y_e) \omega_{|\mathcal{A}|}^{y_e \tilde{y}_e}, \tag{14}$$

where $\omega_{|\mathcal{A}|} = \mathrm{e}^{-2\pi \mathrm{i}/|\mathcal{A}|}$ and $\mathrm{i} = \sqrt{-1}$. See (Bracewell, 1999) for more details on the DFT and (Terras, 1999) for more details on the Fourier transform on finite abelian groups

After applying the dualization procedure and after a little rearranging, we obtain the dual NFG of the 1D Potts model in an external field with periodic boundaries (i.e., the dual of Fig. 1), which is illustrated in Fig. 2.

The big unlabeled boxes in Fig. 2 are the 1D DFT of (10), which can be computed as

$$\tilde{\psi}_e(\tilde{y}_e) = \sum_{y_e \in \mathcal{A}} \mathrm{e}^{\beta J_e \delta(y_e)} \omega_{|\mathcal{A}|}^{y_e \tilde{y}_e} \tag{15}$$

$$= \sum_{y_e \in \mathcal{A}} (1 + (\mathrm{e}^{\beta J_e} - 1)\delta(y_e)) \omega_{|\mathcal{A}|}^{y_e \tilde{y}_e}, \tag{16}$$

which, after a little algebra, gives

$$\tilde{\psi}_e(\tilde{y}_e) = \begin{cases} \mathrm{e}^{\beta J_e} - 1 + q, & \text{if } \tilde{y}_e = 0 \\ \mathrm{e}^{\beta J_e} - 1, & \text{otherwise.} \end{cases} \tag{17}$$

Similarly, the small unlabeled boxes are the 1D DFT of (11) given by

$$\tilde{\phi}_v(\tilde{x}_v) = \begin{cases} \mathrm{e}^{\beta H_v} - 1 + q, & \text{if } \tilde{x}_v = 0 \\ \mathrm{e}^{\beta H_v} - 1, & \text{otherwise.} \end{cases} \tag{18}$$

In the dual NFG, it holds that $\tilde{\mathbf{x}}(\tilde{\mathbf{y}}) = \boldsymbol{M}^{\mathsf{T}} \tilde{\mathbf{y}}$, where $\boldsymbol{M}^{\mathsf{T}}$ denotes the transpose of $\boldsymbol{M}$. In other words, we can freely choose $\tilde{\mathbf{Y}}$ and therefrom compute $\tilde{\mathbf{X}}$. E.g., if we take $\mathcal{G}$ to be a d-dimensional model with pairwise interactions and assume periodic boundary conditions, each component $\tilde{X}_v$ of $\tilde{\mathbf{X}}$ can be computed from the $2d$ components of $\tilde{\mathbf{Y}}$ that are incident to the zero-sum indicator factor attached to $\tilde{X}_v$.

If a ferromagnetic Potts model is under the influence of a nonnegative external field, both factors (17) and (18) will be nonnegative. The global PMF in the dual NFG can thus be defined as

$$\pi_{\mathrm{d}}(\tilde{\mathbf{y}}) = \frac{1}{Z_{\mathrm{d}}} \prod_{e \in \mathcal{E}} \tilde{\psi}_e(\tilde{y}_e) \prod_{v \in \mathcal{V}} \tilde{\phi}_v(\tilde{x}_v), \tag{19}$$

where

$$Z_{\mathrm{d}} = \sum_{\tilde{\mathbf{y}} \in \mathcal{A}^{|\mathcal{E}|}} \prod_{e \in \mathcal{E}} \tilde{\psi}_e(\tilde{y}_e) \prod_{v \in \mathcal{V}} \tilde{\phi}_v(\tilde{x}_v) \tag{20}$$

is the partition function in the dual domain.

In the models that we study in this paper, the NFG duality theorem states that the partition function of the dual NFG Z_{d} is equal to the partition function of the primal NFG Z_{p}, up to some know scale factor. Indeed

$$Z_{\mathrm{d}} = \alpha(\mathcal{G}) Z_{\mathrm{p}} \tag{21}$$

For more details, see (Al-Bashabsheh and Mao, 2011). The scale factor $\alpha(\mathcal{G})$ depends on the topology of $\mathcal{G}$, and is given by

$$\alpha(\mathcal{G}) = |\mathcal{A}|^{|\mathcal{E}| - |\mathcal{V}| + 1} \tag{22}$$

The term $|\mathcal{E}| - |\mathcal{V}| + 1$ that appears in the exponent in (22) is equal to the first Betti number (i.e., the cyclomatic number) of $\mathcal{G}$. See (Molkaraie, 2017, Appendix) and (Forney, 2018) for more details.

The NFG duality theorem as expressed in (21) is closely related to the Holant Theorem (Valiant, 2008). We refer the readers who are interested in this connection to (Cai et al., 2008), (Al-Bashabsheh and Mao, 2011, Section 3.2), and (Cai and Chen, 2017).

4.1 Dual NFG of the Ising Model and the High-Temperature Series Expansion

In the dual NFG of the Ising model

$$\tilde{\psi}_e(\tilde{y}_e) = \begin{cases} 2\cosh(\beta J_e), & \text{if } \tilde{y}_e = 0 \\ 2\sinh(\beta J_e), & \text{if } \tilde{y}_e = 1, \end{cases} \tag{23}$$

which is the 1D DFT of (7). Similarly

$$\tilde{\phi}_v(\tilde{x}_v) = \begin{cases} 2\cosh(\beta H_v), & \text{if } \tilde{x}_v = 0 \\ 2\sinh(\beta H_v), & \text{if } \tilde{x}_v = 1 \end{cases} \tag{24}$$

is the 1D DFT of (8).

As a side remark, note that small and large values of the coupling parameters correspond to the high-temperature and low-temperature regimes, respectively.

Let us consider a ferromagnetic Ising model in a constant and positive external field H. The valid configurations in the dual NFG of this model give rise to the following expansion of the partition function

$$Z_{\mathrm{p}} \propto \sum_{\mathcal{U} \subseteq \mathcal{E}} \tanh(H)^{|\mathrm{odd}(\mathcal{U})|} \prod_{(k,\ell) \in \mathcal{U}} \tanh(J_e), \tag{25}$$

where $\mathcal{U} \subseteq \mathcal{E}$ and $\mathrm{odd}(\mathcal{U})$ denotes the set of all odd-degree vertices in the subgraph of $\mathcal{E}$ induced by $\mathcal{U}$. For more details, see (Molkaraie and Gómez, 2018, Section VIII).

In statistical physics, the sum in (25) is generally known as the high-temperature series expansion of the partition function (Newell and Montroll, 1953) and (Nishimori and Ortiz, 2015, Chapter 10).

In (Jerrum and Sinclair, 1993), the authors have proposed a fully polynomial-time randomized approximation scheme called the subgraphs-world process to estimate the partition function of ferromagnetic Ising models with arbitrary topology and in a positive external field. The subgraphs-world process is based on the high-temperature series expansion of the partition function. Indeed, the configurations in the subgraphs-world process coincide with the configurations in (25).

The mixing time of the subgraphs-world process is polynomial in the size of the model at all temperatures.[2] It is therefore possible to employ the subgraphs-world process to generate configurations in the dual NFG of the ferromagnetic Ising model. Since the process is rapidly mixing, it converges in polynomial time.

5. Marginal Probabilities and Pertinent Mappings

In this section, we derive local mappings that relate marginal probabilities at the edges of a primal NFG to the marginal probabilities at the corresponding edges in the dual NFG. Indeed, we will prove that $(\pi_{\mathrm{p},e}(a)/\psi_e(a), a \in \mathcal{A})$ in the primal domain and $(\pi_{\mathrm{d},e}(a)/\tilde{\psi}_e(a), a \in \mathcal{A})$ in the dual domain are Fourier pairs.

2. This is a remarkable result, as typical Monte Carlo methods (e.g., the Gibbs sampling algorithm for the 2D Ising model) are rapidly mixing only in the high-temperature regime (Martinelli and Olivieri, 1994).

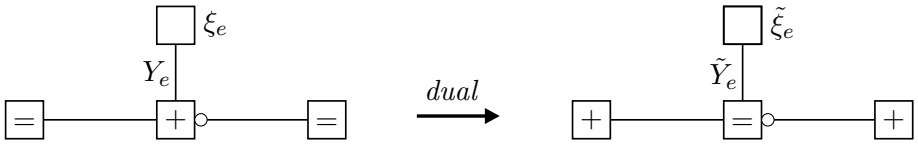

Figure 3: An edge $e \in \mathcal{E}$ in the intermediate primal NFG (left) and in the intermediate dual NFG (right). The unlabeled box (left) represents ξ_e given by (32) and the unlabeled box (right) represents $\tilde{\xi}_e$, the 1D DFT of ξ_e.

In the primal NFG the marginal probability over $e \in \mathcal{E}$ can be computed as

$$\pi_{\mathrm{p},e}(a) = \frac{Z_{\mathrm{p},e}(a)}{Z_{\mathrm{p}}}, \quad a \in \mathcal{A}, \tag{26}$$

where

$$Z_{\mathrm{p},e}(a) = \sum_{\mathbf{x}\colon\, y_e(\mathbf{x})=a} \prod_{e\in\mathcal{E}} \psi_e(y_e(\mathbf{x})) \prod_{v\in\mathcal{V}} \phi_v(x_v) \tag{27}$$

Equivalently,

$$Z_{\mathrm{p},e}(a) = S_e(a)\psi_e(a) \tag{28}$$

with

$$S_e(a) = \sum_{\mathbf{x}\colon\, y_e(\mathbf{x})=a} \prod_{e'\in\mathcal{E}\setminus\{e\}} \psi_{e'}(y_{e'}(\mathbf{x})) \prod_{v\in\mathcal{V}} \phi_v(x_v) \tag{29}$$

The partition function Z_{p} can therefore be written as the dot product of the vectors $S_e(\cdot)$ and $\psi_e(\cdot)$ as

$$Z_{\mathrm{p}} = \langle S_e, \psi_e \rangle \tag{30}$$

$$= \sum_{a\in\mathcal{A}} S_e(a)\psi_e(a), \tag{31}$$

which indicates that (26) is a valid PMF over $\mathcal{A}$.

In Coding Theory terminology, $(\psi_e(a), a \in \mathcal{A})$ is called the *intrinsic* message vector and $(S_e(a), a \in \mathcal{A})$ is called the *extrinsic* message vector at edge $e \in \mathcal{E}$. The fact that at any edge $e \in \mathcal{E}$ the partition function is the dot product of the intrinsic and extrinsic message vectors is called the *sum-product rule*. This observation holds for any edge that is by itself a cut-set of $\mathcal{G}$. For more details, see Forney (2001).

Alternatively, $S_e(a)$ can be viewed as the partition function of an intermediate primal NFG in which all factors are equal to their corresponding factors in the primal NFG, except for $\psi_e(y_e)$, which is replaced by

$$\xi_e(y_e; a) = \delta(y_e - a), \tag{32}$$

as shown in Fig. 3 (left).

The dual of the intermediate primal NFG is shown in Fig. 3 (right), in which, equality indicator factors are replaced by zero-sum indicator factors, and $\xi_e(\cdot)$ is replaced by $\tilde{\xi}_e(\cdot)$,

the 1D DFT of $\xi_e(\cdot)$. We denote the partition function of the dual intermediate NFG by $Z_{\mathrm{d}}^{\mathrm{I}}(a)$. From the NFG duality theorem (21), we obtain

$$Z_{\mathrm{d}}^{\mathrm{I}}(a) = \alpha(\mathcal{G}) S_e(a) \tag{33}$$

Similarly, the edge marginal probability at $e \in \mathcal{E}$ of the dual NFG is given by

$$\pi_{\mathrm{d},e}(a) = \frac{Z_{\mathrm{d},e}(a)}{Z_{\mathrm{d}}}, \quad a \in \mathcal{A}, \tag{34}$$

where

$$Z_{\mathrm{d},e}(a) = \sum_{\tilde{\mathbf{y}}\colon\, \tilde{y}_e = a} \prod_{e \in \mathcal{E}} \tilde{\psi}_e(\tilde{y}_e) \prod_{v \in \mathcal{V}} \tilde{\phi}_v(\tilde{x}_v(\mathbf{y})) \tag{35}$$

The partition function of the dual NFG can be expressed by the following dot product

$$\begin{aligned} Z_{\mathrm{d}} &= \langle T_e, \tilde{\psi}_e \rangle & (36)\\ &= \sum_{a \in \mathcal{A}} T_e(a) \tilde{\psi}_e(a) & (37) \end{aligned}$$

with

$$T_e(a) = \sum_{\tilde{\mathbf{y}}\colon\, \tilde{y}_e = a} \prod_{e' \in \mathcal{E} \setminus \{e\}} \tilde{\psi}_{e'}(\tilde{y}_{e'}) \prod_{v \in \mathcal{V}} \tilde{\phi}_v(\tilde{x}_v(\mathbf{y})) \tag{38}$$

Proposition 1 *The vectors* $(\pi_{\mathrm{p},e}(a)/\psi_e(a), a \in \mathcal{A})$ *and* $(\pi_{\mathrm{d},e}(a)/\tilde{\psi}_e(a), a \in \mathcal{A})$ *are DFT pairs.*

Proof The partition function of the dual intermediate NFG is equal to the dot product of the vectors $(T_e(a), a \in \mathcal{A})$ and $(\tilde{\xi}_e(\tilde{y}_e; a), a \in \mathcal{A})$. From (33), we obtain

$$\begin{aligned} \alpha(\mathcal{G}) S_e(a) &= \langle T_e, \tilde{\xi}_e \rangle & (39)\\ &= \sum_{a' \in \mathcal{A}} T_e(a') \Big(\sum_{y \in \mathcal{A}} \delta(y - a) \omega_{|\mathcal{A}|}^{ya'} \Big) & (40)\\ &= \sum_{a' \in \mathcal{A}} T_e(a') \omega_{|\mathcal{A}|}^{aa'} & (41) \end{aligned}$$

for $a \in \mathcal{A}$. We conclude that the vectors $(S_e(a), a \in \mathcal{A})$ and $(T_e(a), a \in \mathcal{A})$ are Fourier pairs, up to scale factor $\alpha(\mathcal{G})$.

On the other hand (26) and (28) give

$$S_e(a) = Z_{\mathrm{p}} \frac{\pi_{\mathrm{p},e}(a)}{\psi_e(a)}, \quad a \in \mathcal{A}, \tag{42}$$

and combining (34) and (37) yields

$$\begin{aligned} T_e(a) &= Z_{\mathrm{d}} \frac{\pi_{\mathrm{d},e}(a)}{\tilde{\psi}_e(a)} & (43)\\ &= \alpha(\mathcal{G}) Z_{\mathrm{p}} \frac{\pi_{\mathrm{d},e}(a)}{\tilde{\psi}_e(a)}, \quad a \in \mathcal{A}, & (44) \end{aligned}$$

where (44) follows from the NFG duality theorem in (21).

Substituting $(S_e(a), a \in \mathcal{A})$ in (42) and $(T_e(a), a \in \mathcal{A})$ in (44) into (41) yields

$$\frac{\pi_{\mathrm{p},e}(a)}{\psi_e(a)} = \sum_{a' \in \mathcal{A}} \frac{\pi_{\mathrm{d},e}(a')}{\tilde{\psi}_e(a')} \omega_{|\mathcal{A}|}^{aa'}, \tag{45}$$

which completes the proof. ■

We state without proof that

Proposition 2 *The vectors* $(\pi_{\mathrm{p},v}(a)/\phi_v(a), a \in \mathcal{A})$ *and* $(\pi_{\mathrm{d},v}(a)/\tilde{\phi}_v(a), a \in \mathcal{A})$ *are DFT pairs.*

By virtue of Propositions 1 and 2, it is possible to estimate the edge marginal probabilities in one domain, and then transform them to the other domain all together. In particular, for the ferromagnetic Ising model in a positive external field we can employ the subgraphs-world process (as a rapidly mixing Markov chain) to generate configuration in the dual NFG of the model.

We finish this section with the following important remarks.

Remark 3 *The proposed mappings are fully local (i.e., edge-dependent or vertex-dependent). Moreover, the mappings do not depend on the size and on the topology of* $\mathcal{G}$*. Indeed, all the relevant information regarding the rest of the graph is already incorporated in the estimated edge marginal densities.*

Remark 4 *Transforming marginals from one domain to the other requires computing a DFT with computational complexity* $\mathcal{O}(|\mathcal{A}|^2)$*, which can be reduced to* $\mathcal{O}(|\mathcal{A}| \log(|\mathcal{A}|))$ *via the fast Fourier transform (FFT). However, when there is symmetry in the factors, as in the case of the Ising and the Potts models, the complexity can be further reduced to* $\mathcal{O}(|\mathcal{A}|)$*.*

Remark 5 *In binary models, the factors in the dual NFG can in general take negative values, and in nonbinary models, the factors can even be complex-valued. In such cases a valid PMF can no longer be defined in the dual domain.The mapping nevertheless remains valid; but for marginal functions, instead of marginal densities, of a global function with a factorization given by (19).*

6. Details of the Mapping for Binary Models

In binary models (i.e., when $\mathcal{A} = \mathbb{Z}/2\mathbb{Z}$), Proposition 1 provides the following mapping

$$\begin{pmatrix} \pi_{\mathrm{p},e}(0)/\psi_e(0) \\ \pi_{\mathrm{p},e}(1)/\psi_e(1) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \pi_{\mathrm{d},e}(0)/\tilde{\psi}_e(0) \\ \pi_{\mathrm{d},e}(1)/\tilde{\psi}_e(1) \end{pmatrix} \tag{46}$$

in matrix-vector format via the two-point DFT matrix, where

$$\tilde{\psi}_e(0) = \psi_e(0) + \psi_e(1) \tag{47}$$

and

$$\tilde{\psi}_e(1) = \psi_e(0) - \psi_e(1) \tag{48}$$

For a general Ising model, substituting factors (7) and (23) in (46) yields

$$\begin{pmatrix} \pi_{\mathrm{p},e}(0) \\ \pi_{\mathrm{p},e}(1) \end{pmatrix} = \begin{pmatrix} \dfrac{\mathrm{e}^{\beta J_e}}{2\cosh(\beta J_e)} & \dfrac{\mathrm{e}^{\beta J_e}}{2\sinh(\beta J_e)} \\ \dfrac{\mathrm{e}^{-\beta J_e}}{2\cosh(\beta J_e)} & -\dfrac{\mathrm{e}^{-\beta J_e}}{2\sinh(\beta J_e)} \end{pmatrix} \begin{pmatrix} \pi_{\mathrm{d},e}(0) \\ \pi_{\mathrm{d},e}(1) \end{pmatrix} \tag{49}$$

for $\beta J_e \neq 0$, which implies

$$\begin{pmatrix} \pi_{\mathrm{d},e}(0) \\ \pi_{\mathrm{d},e}(1) \end{pmatrix} = \begin{pmatrix} \dfrac{\cosh(\beta J_e)}{\mathrm{e}^{\beta J_e}} & \dfrac{\cosh(\beta J_e)}{\mathrm{e}^{-\beta J_e}} \\ \dfrac{\sinh(\beta J_e)}{\mathrm{e}^{\beta J_e}} & -\dfrac{\sinh(\beta J_e)}{\mathrm{e}^{-\beta J_e}} \end{pmatrix} \begin{pmatrix} \pi_{\mathrm{p},e}(0) \\ \pi_{\mathrm{p},e}(1) \end{pmatrix} \tag{50}$$

As a sanity check, we consider a ferromagnetic and homogeneous Ising model in the low-temperature and high-temperature limits. In the primal NFG and in the low-temperature limit (i.e., as $\beta J_e \to +\infty$), the edge-marginal probability $\pi_{\mathrm{p},e}(0) = 1 - \pi_{\mathrm{p},e}(1) = 1$, which is in agreement with (49). In the high-temperature limit (i.e., as $\beta J_e \to 0$), the marginal probability over edge e of the dual NFG is $\pi_{\mathrm{d},e}(0) = 1 - \pi_{\mathrm{d},e}(1) = 1$. This is in agreement with (50). (Recall that β is proportional to the inverse temperature.)

Proposition 6 *In an arbitrary ferromagnetic Ising model in a nonnegative external field, it holds that*

$$\pi_{\mathrm{p},e}(0) \geq \frac{1}{1 + \mathrm{e}^{-2\beta J_e}} \tag{51}$$

and

$$\pi_{\mathrm{d},e}(0) \geq \frac{1 + \mathrm{e}^{-2\beta J_e}}{2} \tag{52}$$

Proof In the dual domain, the global PMF of a ferromagnetic Ising model in a nonnegative field $\pi_{\mathrm{d},e}(\cdot)$ is given by (19). From (49)

$$\frac{\pi_{\mathrm{p},e}(0)}{\mathrm{e}^{\beta J_e}} = \frac{\pi_{\mathrm{d},e}(0)}{2\cosh(\beta J_e)} + \frac{\pi_{\mathrm{d},e}(1)}{2\sinh(\beta J_e)} \tag{53}$$

$$= \frac{1}{2\sinh(\beta J_e)} - \frac{\mathrm{e}^{-\beta J_e}}{\sinh(2\beta J_e)} \pi_{\mathrm{d},e}(0) \tag{54}$$

Therefore $\pi_{\mathrm{p},e}(0)$ achieves its minimum when $\pi_{\mathrm{d},e}(0) = 1$. After substituting $\pi_{\mathrm{d},e}(0) = 1$ in (54), and after a little algebra, we obtain

$$\pi_{\mathrm{p},e}(0) \geq \frac{\mathrm{e}^{\beta J_e}}{2\cosh(\beta J_e)} \tag{55}$$

$$= \frac{1}{1 + \mathrm{e}^{-2\beta J_e}} \tag{56}$$

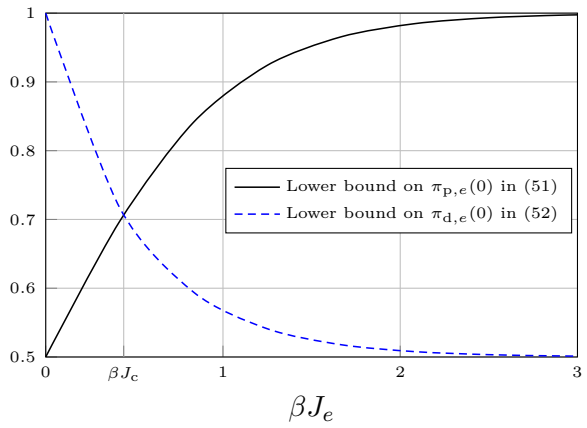

Figure 4: For a ferromagnetic Ising model in a nonnegative external field, the solid black plot and the dashed blue plot show the lower bound on $\pi_{\mathrm{p},e}(0)$ given by (51), and the lower bound on $\pi_{\mathrm{d},e}(0)$ given by (52), as a function of βJ_e, respectively. The lower bounds intersect at the criticality of the 2D homogeneous Ising model in the absence of an external field, denoted by βJ_{c}.

The proof of (52) follows along the same lines by starting from (50), and is omitted. ■

From (51) and (52), we conclude that in an arbitrary ferromagnetic Ising model in a nonnegative external field

$$\pi_{\mathrm{p},e}(0)\pi_{\mathrm{d},e}(0) \geq \frac{1}{2}, \tag{57}$$

which is in the form of an uncertainty principle.

Fig. 4 shows the lower bounds in (51) and in (52) as a function of βJ_e. Interestingly, the lower bounds intersect at $\beta J_c = \ln(1+\sqrt{2})/2$, which coincides with the critical coupling (i.e., the phase transition) of the 2D homogeneous Ising model in zero field and in the thermodynamic limit (i.e., as $|\mathcal{V}| \to \infty$). For more details, see (Kramers and Wannier, 1941; Onsager, 1944).

We stress out that the lower bonds in Proposition 6 are valid for arbitrary ferromagnetic Ising models in a nonnegative external field. The bounds do not depend on the size or on the topology of $\mathcal{G}$. In Appendix B, we will prove that the edge marginal probability of the 1D Ising model with free boundary conditions attains the lower bound in (51).

6.1 The Fixed Points

Let us denote the fixed points of the mapping in (46) by $(\pi_e^*(a), a \in \mathcal{A})$. A routine calculation gives

$$\begin{pmatrix} \pi_e^*(0) & \pi_e^*(1) \end{pmatrix} = \begin{pmatrix} \dfrac{\psi_e(0)\tilde{\psi}_e(0)}{\psi_e(0)\tilde{\psi}_e(0)+\psi_e(1)\tilde{\psi}_e(1)} & \dfrac{\psi_e(1)\tilde{\psi}_e(1)}{\psi_e(0)\tilde{\psi}_e(0)+\psi_e(1)\tilde{\psi}_e(1)} \end{pmatrix} \tag{58}$$

For a homogeneous and ferromagnetic Ising model we thus obtain

$$\begin{pmatrix} \pi^*(0) & \pi^*(1) \end{pmatrix} = \begin{pmatrix} \dfrac{\mathrm{e}^{\beta J}\cosh(\beta J)}{1+\sinh(2\beta J)} & \dfrac{\mathrm{e}^{-\beta J}\sinh(\beta J)}{1+\sinh(2\beta J)} \end{pmatrix} \tag{59}$$

Fig. 5 shows the fixed points $\pi^*(\cdot)$ as a function of βJ.

Proposition 7 *The minimum of $\pi^*(0)$ and the maximum of $\pi^*(1)$ are attained at the criticality of the 2D homogeneous Ising model in the absence of an external magnetic field.*

Proof From (6), the Hamiltonian of the model is

$$\mathcal{H}(\mathbf{y}) = -\sum_{e\in\mathcal{E}} J\big(2\delta(y_e)-1\big) \tag{60}$$

$$= -J\sum_{e\in\mathcal{E}}(1-2y_e) \tag{61}$$

Consequently, the average energy can be computed as

$$\overline{\mathcal{H}}(\mathbf{y}) = \sum_{\mathbf{y}\in\mathcal{A}} \pi_{\mathrm{p}}(\mathbf{y})\mathcal{H}(\mathbf{y}) \tag{62}$$

$$= -J|\mathcal{E}|(1-2\,\mathrm{E}[Y_e]) \tag{63}$$

$$= -J|\mathcal{E}|(1-2\pi_{\mathrm{p},e}(1)) \tag{64}$$

In a 2D Ising model with periodic boundaries $|\mathcal{E}| = 2|\mathcal{V}|$, therefore the average energy per site is equal to

$$\overline{\mathcal{H}}(\mathbf{y})/|\mathcal{V}| = -2J(1-2\pi_{\mathrm{p},e}(1)) \tag{65}$$

From Onsager's closed-form solution, in the thermodynamic limit, the logarithm of the partition function (i.e., the free energy) per site of the 2D homogeneous Ising model in zero field is given by

$$\lim_{|\mathcal{V}|\to\infty}\frac{\ln Z(\beta J)}{|\mathcal{V}|} = \frac{1}{2}\ln(2\cosh^2 2\beta J) + \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \ln\Big(1+(1-\kappa^2\sin^2\theta)^{1/2}\Big)\,d\theta \tag{66}$$

and the average (internal) energy per site $U(\beta J)$ is

$$U(\beta J) = -\lim_{|\mathcal{V}|\to\infty}\frac{1}{|\mathcal{V}|}\cdot\frac{\partial \ln Z(\beta J)}{\partial\beta} \tag{67}$$

$$= -J\coth(2\beta J)\Big(1-\frac{1}{2\pi}(1-\kappa\sinh 2\beta J)\int_0^{\frac{\pi}{2}}(1-\kappa^2\sin^2\theta)^{-1/2}d\theta\Big) \tag{68}$$

with

$$\kappa(\beta J) = \frac{2\sinh 2\beta J}{\cosh^2 2\beta J} \tag{69}$$

For more details, see (Onsager, 1944) and (Baxter, 2007, Chapter 7).

At criticality (i.e., at $\beta J_{\mathrm{c}} = \ln(1+\sqrt{2})/2$) it is easy to show that $\kappa(\beta J_{\mathrm{c}}) = 1$ and the average energy per site $U(\beta J_{\mathrm{c}})$ is

$$U(\beta J_{\mathrm{c}}) = -\sqrt{2}J_{\mathrm{c}} \tag{70}$$

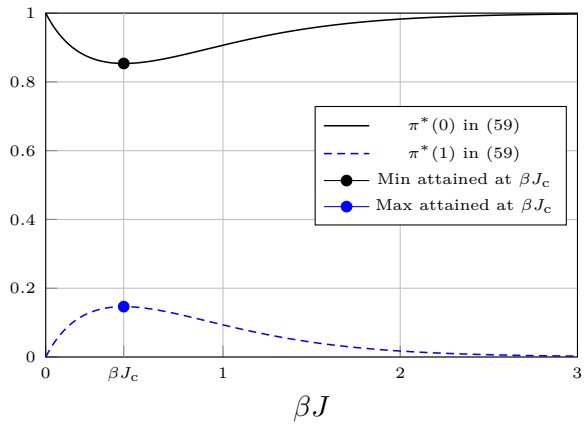

Figure 5: The fixed points (59) as a function of βJ. The filled circles show the fixed points at the criticality of the 2D homogeneous Ising model in zero field as in (71).

From (65) and (70), we conclude that in the thermodynamic limit

$$\Big(\pi(0) \quad \pi(1)\Big)\Big|_{\beta J=\beta J_c} = \Big((2+\sqrt{2})/4 \quad (2-\sqrt{2})/4\Big)\,, \tag{71}$$

which coincides with the minimum of $\pi^*(0)$ and the maximum of $\pi^*(1)$ in (59). ■

The fixed points at βJ_c are illustrated by filled circles in Fig. 5. We emphasize that, in the thermodynamic limit and in the absence of an external field, the edge marginal probabilities in the primal and in the dual of the 2D homogeneous Ising model are equal at criticality (i.e., at the phase transition).

As side remark, we point out that the 2D Ising model and its criticality have been studies in the context of generative neural networks and restricted Boltzmann machines, see (Morningstar and Melko, 2018) and (Cossu et al., 2019).

6.2 The Local Magnetization

In a ferromagnetic Ising model, let

$$\Delta_{\mathrm{p},e} = \pi_{\mathrm{p},e}(0) - \pi_{\mathrm{p},e}(1) \tag{72}$$

and

$$\Delta_{\mathrm{d},e} = \pi_{\mathrm{d},e}(0) - \pi_{\mathrm{d},e}(1) \tag{73}$$

denote the local magnetizations at edge $e \in \mathcal{E}$ of the primal NFG and the dual NFG, respectively. Clearly, it is possible to compute $(\pi_{\mathrm{p},e}(a), a \in \mathcal{A})$ from $\Delta_{\mathrm{p},e}$ and $(\pi_{\mathrm{d},e}(a), a \in \mathcal{A})$ from $\Delta_{\mathrm{d},e}$. Furthermore, from the mappings in (49) and (50) we get

$$\pi_{\mathrm{p},e}(0) = \frac{\mathrm{e}^{2\beta J_e} - \Delta_{\mathrm{d},e}}{2\sinh(2\beta J_e)} \tag{74}$$

and

$$\pi_{\mathrm{d},e}(0) = \frac{\coth(\beta J_e) - \Delta_{\mathrm{p},e}}{2\,\mathrm{csch}(2\beta J_e)}, \tag{75}$$

where $\mathrm{csch}(\cdot) = 1/\sinh(\cdot)$.

From (72) and (74) it is easy to verify that

$$\Delta_{\mathrm{p},e} = \frac{\cosh(2\beta J_e) - \Delta_{\mathrm{d},e}}{\sinh(2\beta J_e)}, \tag{76}$$

which relates the local magnetizations in the primal and in the dual domains. The local magnetizations $\Delta_{\mathrm{p},e}$ and $\Delta_{\mathrm{d},e}$ are both equal to $\sqrt{2}/2$ at βJ_{c} (i.e., at the criticality of the 2D homogeneous Ising model).

7. Details of the Mapping for Non-binary Models

In non-binary models (e.g., the Potts and the clock models) the mapping between the vecotors $(\pi_{\mathrm{p},e}(a)/\psi_e(a), a \in \mathcal{A})$ and $(\pi_{\mathrm{d},e}(a)/\tilde{\psi}_e(a), a \in \mathcal{A})$ is given by

$$\left(\pi_{\mathrm{p},e}(a)/\psi_e(a), a \in \mathcal{A}\right) = \boldsymbol{W}_{|\mathcal{A}|\times|\mathcal{A}|} \left(\pi_{\mathrm{d},e}(a)/\tilde{\psi}_e(a), a \in \mathcal{A}\right), \tag{77}$$

where $(\pi_{\mathrm{p},e}(a)/\psi_e(a), a \in \mathcal{A})$ and $(\pi_{\mathrm{d},e}(a)/\tilde{\psi}_e(a), a \in \mathcal{A})$ are column vectors of length $|\mathcal{A}|$, and $\boldsymbol{W}_{|\mathcal{A}|\times|\mathcal{A}|}$ is the $|\mathcal{A}|$-point DFT matrix (i.e., the Vandermonde matrix for the roots of unity), in which

$$\boldsymbol{W}_{k,\ell} = \omega_{|\mathcal{A}|}^{k\ell}, \quad k, \ell \in \mathcal{A}, \tag{78}$$

where $\omega_{|\mathcal{A}|} = \mathrm{e}^{-2\pi\mathrm{i}/|\mathcal{A}|}$.

E.g., in a three-state Potts model the mapping boils down to

$$\begin{pmatrix} \pi_{\mathrm{p},e}(a)/\psi_e(0) \\ \pi_{\mathrm{p},e}(a)/\psi_e(1) \\ \pi_{\mathrm{p},e}(a)/\psi_e(2) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega^4 \end{pmatrix} \begin{pmatrix} \pi_{\mathrm{d},e}(0)/\tilde{\psi}_e(0) \\ \pi_{\mathrm{d},e}(1)/\tilde{\psi}_e(1) \\ \pi_{\mathrm{d},e}(2)/\tilde{\psi}_e(2) \end{pmatrix} \tag{79}$$

for $\beta J_e \neq 0$, which implies

$$\begin{pmatrix} \pi_{\mathrm{d},e}(0)/\tilde{\psi}_e(0) \\ \pi_{\mathrm{d},e}(1)/\tilde{\psi}_e(1) \\ \pi_{\mathrm{d},e}(2)/\tilde{\psi}_e(2) \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega^{-1} & \omega^{-2} \\ 1 & \omega^{-2} & \omega^{-4} \end{pmatrix} \begin{pmatrix} \pi_{\mathrm{p},e}(a)/\psi_e(0) \\ \pi_{\mathrm{p},e}(a)/\psi_e(1) \\ \pi_{\mathrm{p},e}(a)/\psi_e(2) \end{pmatrix} \tag{80}$$

with $\omega = \mathrm{e}^{-2\pi\mathrm{i}/3}$.

Due to symmetry in factors (10) in the primal NFG and factors (17) in the dual NFG of the q-state Potts model, it holds that

$$\frac{\pi_{\mathrm{p},e}(1)}{\psi_e(1)} = \frac{\pi_{\mathrm{p},e}(2)}{\psi_e(2)} = \cdots = \frac{\pi_{\mathrm{p},e}(q-1)}{\psi_e(q-1)} \tag{81}$$

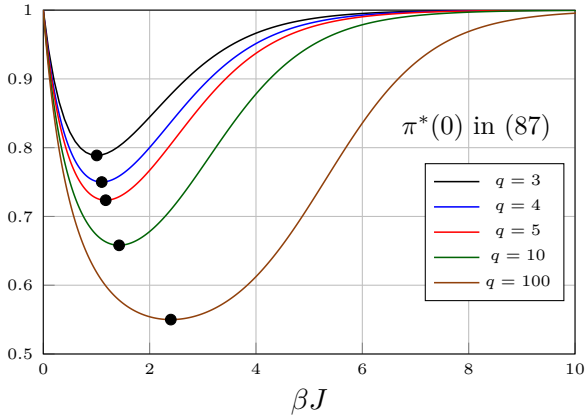

Figure 6: The fixed points (87) as a function of βJ for different values of q. The filled circles show the fixed points at the criticality of the 2D homogeneous Potts model in zero field located at $\beta J_{\rm c} = \ln(1+\sqrt{q})$.

and

$$\frac{\pi_{{\rm d},e}(1)}{\tilde{\psi}_e(1)} = \frac{\pi_{{\rm d},e}(2)}{\tilde{\psi}_e(2)} = \cdots = \frac{\pi_{{\rm d},e}(q-1)}{\tilde{\psi}_e(q-1)} \tag{82}$$

Hence

$$\frac{\pi_{{\rm p},e}(0)}{\psi_e(0)} = \sum_{a\in\mathcal{A}} \frac{\pi_{{\rm d},e}(a)}{\tilde{\psi}_e(a)} \tag{83}$$

and for $t \in \{1, 2, \ldots, q-1\}$

$$\frac{\pi_{{\rm p},e}(t)}{\psi_e(t)} = \sum_{a\in\mathcal{A}} \frac{\pi_{{\rm d},e}(a)}{\tilde{\psi}_e(a)} \omega^{at}_{|\mathcal{A}|} \tag{84}$$

$$= \frac{\pi_{{\rm d},e}(0)}{\tilde{\psi}_e(0)} - \frac{\pi_{{\rm d},e}(1)}{\tilde{\psi}_e(1)} \tag{85}$$

which are real-valued.

A straightforward generalization of (58) gives the fixed points $(\pi^*_e(a), a \in \mathcal{A})$ of the mapping in (77) as

$$\pi^*_e(a) = \frac{\psi_e(a)\tilde{\psi}_e(a)}{S}, \quad a \in \mathcal{A} \tag{86}$$

where $S = \sum_{a\in\mathcal{A}} \psi_e(a)\tilde{\psi}_e(a)$.

For a homogeneous and ferromagnetic Potts model, the fixed points are given by

$$\pi^*(0) = \frac{{\rm e}^{\beta J}({\rm e}^{\beta J} - 1 + q)}{{\rm e}^{2\beta J} - 2(1-q){\rm e}^{\beta J} + 1 - q} \tag{87}$$

and

$$\pi^*(t) = \frac{{\rm e}^{\beta J} - 1}{{\rm e}^{2\beta J} - 2(1-q){\rm e}^{\beta J} + 1 - q} \tag{88}$$

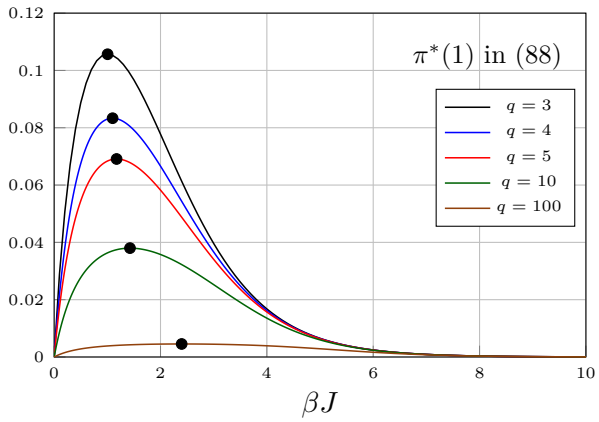

Figure 7: The fixed points (88) as a function of βJ for different values of q. The filled circles show the fixed points at the criticality of the 2D homogeneous Potts model in zero field located at $\beta J_{\rm c} = \ln(1+\sqrt{q})$.

for $t \in \{1, 2, \ldots, q-1\}$.

Fig. 6 shows the fixed point (87) as a function of βJ for different values of q. As in the case of the Ising model, the minimum of $\pi^*(0)$ is attained at the criticality of the 2D homogeneous Potts model without an external field, which is located at $\beta J_{\rm c} = \ln(1+\sqrt{q})$. The fixed points in (88) are plotted in Fig. 7. We observe that the maximum of $(\pi^*(t), t \in \{1, 2, \ldots, q-1\})$ are also attained at the criticality. For more details on the phase transition of the 2D Potts model, see (Potts, 1952; Wu, 1982).

A closed-form solution for the partition function of the 2D Potts model is not available. However, from (87) and (88) we can still obtain the values of the fixed points at criticality, which are given by

$$\pi^*(0) = \frac{1}{2}(1 + \frac{1}{\sqrt{q}}) \tag{89}$$

and

$$\pi^*(t) = \frac{1}{2(q-1)}(1 - \frac{1}{\sqrt{q}}), \tag{90}$$

for $t \in \{1, 2, \ldots, q-1\}$. The filled circles in Figs. 6 and 7 show the fixed points at criticality of the 2D homogeneous Potts model for different values of q.

In summary

$$\Big(\pi^*(0) \quad \pi^*(1) \quad \pi^*(2) \quad \cdots \quad \pi^*(q-1)\Big)\Big|_{\beta J = \beta J_{\rm c}} = \frac{1}{2\sqrt{q}}\begin{pmatrix} 1+\sqrt{q} & \frac{1}{1+\sqrt{q}} & \frac{1}{1+\sqrt{q}} & \cdots & \frac{1}{1+\sqrt{q}} \end{pmatrix} \tag{91}$$

In the many-component limit (i.e., as $q \to \infty$), we obtain

$$\lim_{q\to\infty} \pi^*(0) = \frac{1}{2} \tag{92}$$

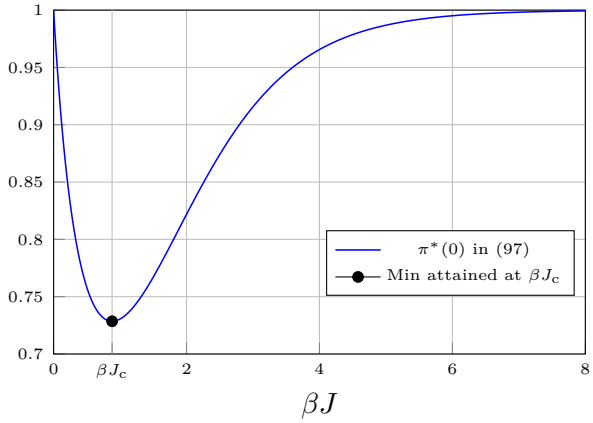

Figure 8: The fixed points of $\pi^*(0)$ in (97) as a function of βJ. The filled circle shows the fixed point at the criticality of the 2D homogeneous four-state clock model, which is located at $\beta J_c = \ln(1+\sqrt{2})$.

We state without proof that

Proposition 8 *In an arbitrary ferromagnetic Potts model in a nonnegative external field, it holds that*

$$\pi_{\mathrm{p},e}(0) \geq \frac{\mathrm{e}^{\beta J_e}}{\mathrm{e}^{\beta J_e} - 1 + q} \tag{93}$$

and

$$\pi_{\mathrm{d},e}(0) \geq \frac{\mathrm{e}^{\beta J_e} - 1 + q}{q\mathrm{e}^{\beta J_e}}, \tag{94}$$

which gives

$$\pi_{\mathrm{p},e}(0)\pi_{\mathrm{d},e}(0) \geq \frac{1}{q} \tag{95}$$

As in the case of the Ising model, the bounds in (93) and (94) intersect at the criticality of the 2D homogenous Potts model, cf. Fig. 4

We briefly mention analogous results for the 2D homogeneous four-state clock model in zero field, which has a phase transition at $\beta J_c = \ln(1+\sqrt{2})$, see (Kihara et al., 1954).

In the clock model

$$\tilde{\psi}(\tilde{y}_e) = \begin{cases} 2(\cosh(\beta J)+1), & \text{if } \tilde{y}_e = 0 \\ 2(\cosh(\beta J)-1), & \text{if } \tilde{y}_e = 2 \\ 2\sinh(\beta J), & \text{otherwise,} \end{cases} \tag{96}$$

which is the 1D DFT of (13), and is positive if the model is ferromagnetic (i.e., if $\beta J > 0$).

For this model, the fixed points of the mapping (77) at criticality are as in

$$\begin{pmatrix} \pi^*(0) & \pi^*(1) & \pi^*(2) & \pi^*(3) \end{pmatrix}\Big|_{\beta J = \beta J_c} = \\ \frac{1}{4(1+\sinh(\beta J))^2} \begin{pmatrix} (e^{\beta J}+1)^2 & 2\sinh(\beta J) & (e^{-\beta J}-1)^2 & 2\sinh(\beta J) \end{pmatrix}, \tag{97}$$

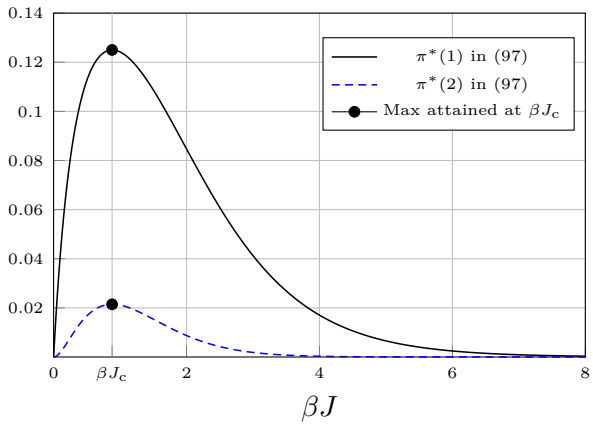

Figure 9: The fixed points of $\pi^*_{\mathrm{p},e}(1)$ and $\pi^*_{\mathrm{p},e}(2)$ in (97) as a function of βJ. The filled circles show the fixed points at the criticality of the 2D homogeneous four-state clock model located at $\beta J_{\mathrm{c}} = \ln(1+\sqrt{2})$.

which are plotted as a function of βJ in Figs. 8 and 9.

The filled circles show the fixed points at the criticality of the model, given by

$$\Big(\pi^*(0) \quad \pi^*(1) \quad \pi^*(2) \quad \pi^*(3)\Big)\Big|_{\beta J=\beta J_{\mathrm{c}}} = \left(\frac{3+2\sqrt{2}}{8} \quad \frac{1}{8} \quad \frac{3-2\sqrt{2}}{8} \quad \frac{1}{8}\right), \tag{98}$$

which coincide with the minimum of $\pi^*(0)$ and the maximums of $\pi^*(1)$ and $\pi^*(2)$ in (97).

The clock model exhibits Kosterlitz-Thouless transitions for large values of q, which is beyond the scope of this paper. We refer interested readers to (Kosterlitz and Thouless, 1973) and (Nishimori and Ortiz, 2015, Chapter 7).

8. Numerical Experiments

In both domains, estimates of marginal densities can be obtained via Markov chain Monte Carlo methods or via variational algorithms (Robert and Casella, 2004; Murphy, 2012). We only consider the subgraphs-world process (SWP) and two variational algorithms, the BP and the TEP algorithms, for ferromagnetic Ising models and frustrated Potts models. Estimated marginals in the dual domain are then transformed all together to the primal domain via (49). In all experiments, the exact values of marginal probabilities are computed via the junction tree algorithm implemented in (Mooij, 2010).

In our first experiment, we consider a 2D homogeneous Ising model, in a constant external field $\beta H = 0.15$, with periodic boundaries, and with size $N = 6 \times 6$. For this model, BP and TEP in the primal and in the dual domains give virtually indistinguishable approximations. We also apply SWP using 10^5 samples. Fig. 10 shows the relative error in estimating $\pi_{\mathrm{p},e}(0)$ as a function of βJ.

We observe that SWP provides good estimates of the marginal probability in the whole range. However, compared to variational algorithms, convergence of the SWP is slow; moreover, SWP is only applicable when the external field is nonzero. In the next two

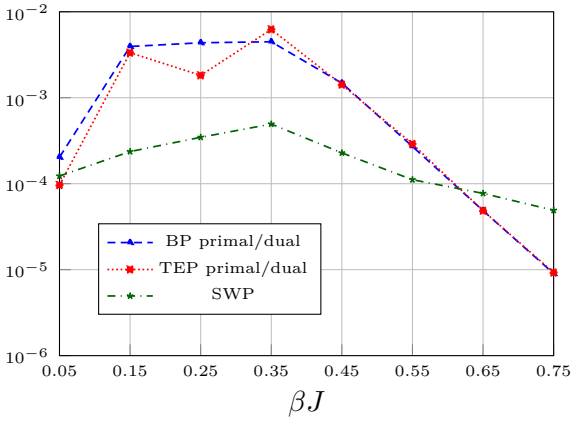

Figure 10: Relative error as a function of βJ in estimating $\pi_{\mathrm{p},e}(0)$ of a homogeneous Ising model in a constant external field $\beta H = 0.15$, with periodic boundaries, and with size $N = 6 \times 6$.

experiments, we consider Ising models in the absence of an external field, and only compare the efficiency of variational algorithms employed in the primal and in the dual domains.

In the second experiment, we consider a 2D ferromagnetic Ising model with size $N = 6 \times 6$, in zero field, and with periodic boundaries. Couplings are chosen randomly according to a half-normal distribution, i.e., $\beta J_e = |\beta J'_e|$ with $\beta J'_e \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$.

Fig. 11 shows the average relative error in estimating the marginal probability $\pi_{\mathrm{p},e}(0)$ as a function of σ^2, where the results are averaged over 200 independent realizations. We consider a fully-connected Ising model with $N = 10$ in our third experiment. Couplings are chosen randomly according to $\beta J_e \overset{\text{i.i.d.}}{\sim} \mathcal{U}[0.05, \beta J_x]$, i.e., uniformly between 0.05 and βJ_x denoted by the value on the x-axis. The average relative error over 50 independent realizations is illustrated in Fig. 12.

In the second and the third experiments, BP and TEP provide close approximations in the dual domain, therefore only BP results are reported. Figs. 11 and 12 show that for $\sigma^2 > 0.25$ and $\beta J_x > 0.20$ (i.e., in relatively lower temeratures), BP in the dual NFG can significantly improve the quality of estimates – even by more than two orders of magnitude in terms of relative error.

In our last experiment, we consider a 2D 3-state Potts model with size $N = 6 \times 6$, in the absence of an external field, and with free boundary conditions, in which all plaquettes (i.e., cycles of length four) are frustrated.

A plaquette is called frustrated if the product of four coupling parameters along its edges is negative. It is then not possible to satisfy all local constraints at the same time, which leads to difficult energy landscapes (Nishimori and Ortiz, 2015). We then focus on the plaquette in the middle of the model.

In order to create frustration, each plaquette has one coupling parameter set to $\beta J_{e_{\text{Antif}}} = -0.25$ (i.e., with antiferromagnetic interaction), and three remaining couplings equal to $\beta J_{e_{\text{Ferr}}}$ according to the value on the x-axis (i.e., with ferromagnetic interaction). According

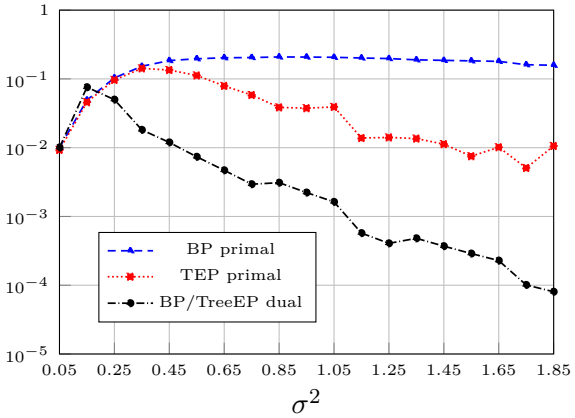

Figure 11: Average relative error in estimating $\pi_{\mathrm{p},e}(0)$ of an Ising model with periodic boundaries and with size $N = 6 \times 6$. Couplings are chosen randomly according to a half-normal distribution with variance σ^2.

to (17), in this example factors with antiferromagnetic interactions will have negative components in the dual NFG, c.f. Remark 4.

Fig. 13 shows the relative error in estimating the edge marginal function of an edge with ferromagnetic interaction as a function of $\beta J_{e_{\mathrm{Ferr}}}$. In this example, BP in the dual domain provides the most accurate estimates when $\beta J_{e_{\mathrm{Ferr}}} > 1.90$. For smaller values of $\beta J_{e_{\mathrm{Ferr}}}$, BP in the primal and BP in the dual domains perform similarly, and for $\beta J_{e_{\mathrm{Ferr}}} < 1.0$, TEP in the primal domain gives the best estimates.

In general, we observe that applying inference algorithms on the dual NFG is advantageous when the coupling parameters are large (i.e., when the temperature is low).

9. Extensions to Continuous Models

As an extension of the proposed mappings to continuous models, we consider the following probability density function (PDF) in the primal domain

$$f_{\mathrm{p}}(\mathbf{x}) \propto \exp\Big(-\frac{1}{2s^2}\sum_{(k,\ell)\in\mathcal{E}}(x_k - x_\ell)^2\Big)\exp\Big(-\frac{1}{2\sigma^2}\sum_{v\in\mathcal{V}}x_v^2\Big) \tag{99}$$

$$= \exp\Big(-\frac{1}{2s^2}\sum_{e\in\mathcal{E}}y_e^2\Big)\exp\Big(-\frac{1}{2\sigma^2}\sum_{v\in\mathcal{V}}x_v^2\Big), \tag{100}$$

where $\mathbf{x} \in \mathbb{R}^{|\mathcal{V}|}$, s^2 denote the intervariable variance, and σ^2 denotes the vertex variance. This Gaussian Markov random field with the thin-membrane prior is widely used in Bayesian image analysis, in data interpolation, and in learning the structure of Gaussian graphical models (Winkler, 1995; Weiss and Freeman, 2000; Malioutov, 2008; Yu et al., 2022).

The PDF in (100) is in accordance with the formulation in Section 2, as the edge-weighing factor $\psi_e(\cdot)$ is only a function of the edge configuration y_e. Indeed from (4), we note that

$$\psi_e(y_e) = \exp(-\frac{1}{2s^2}y_e^2) \tag{101}$$

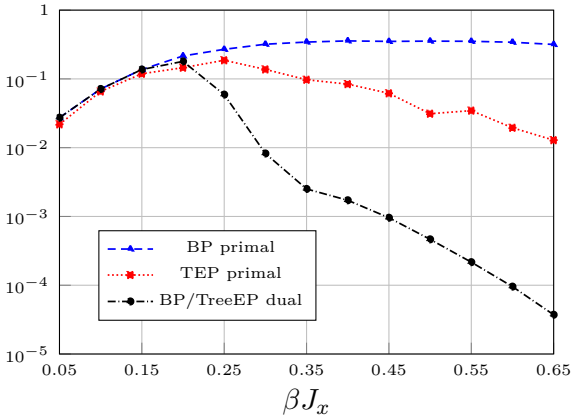

Figure 12: Average relative error in estimating $\pi_{\mathrm{p},e}(0)$ in a fully-connected Ising model with $N = 10$. Coupling parameters are chosen uniformly and independently between 0.05 and βJ_x denoted by the x-axis.

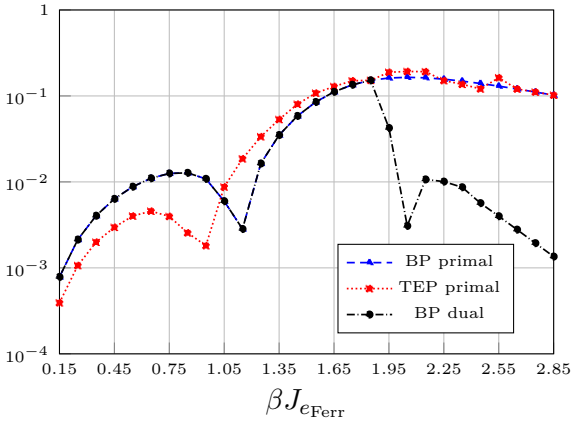

Figure 13: Relative error in estimating the edge marginal function of an edge with ferromagnetic interaction of a frustrated 3-state Potts model with free boundaries and size $N = 6 \times 6$. Here e_{Ferr} has a ferromagnetic interaction $\beta J_{e_{\mathrm{Ferr}}}$.

and the vertex-weighing factors are given by

$$\phi_v(x_v) = \exp\big(-\frac{1}{2\sigma^2}x_v^2\big) \tag{102}$$

For a chain graph, the primal NFG of (100) is illustrated in Fig. 1, where the big unlabeled boxes represent (101), the small unlabeled boxes represent (102), boxes labeled "=" are instances of equality indicator factors $\delta_{=}(\cdot)$, where $\delta_{=}(\cdot)$ denotes the Dirac delta function, and boxes labeled "+" are instances of zero-sum indicator factors, which impose the constraint that all their incident variables sum to zero in $\mathbb{R}$.

In this framework $(\mathcal{F}\psi_e)(\cdot)$, the 1D Fourier transform of $\psi_e(\cdot)$, is the function $\tilde{\psi}_e(\cdot)$ which is given by

$$\tilde{\psi}_e(\tilde{y}_e) = (\mathcal{F}\psi_e)(\tilde{y}_e) \tag{103}$$

$$= \int_{-\infty}^{\infty} \psi_e(y_e) \mathrm{e}^{-\mathrm{i}\,\tilde{y}_e y_e} dy_e \tag{104}$$

where i is the unit imaginary number (Bracewell, 1999).

In analogy with (19), the PDF in the dual domain can be expressed as

$$f_{\mathrm{d}}(\tilde{\mathbf{y}}) \propto \prod_{e\in\mathcal{E}} \tilde{\psi}_e(\tilde{y}_e) \prod_{v\in\mathcal{V}} \tilde{\phi}_v(\tilde{x}_v), \tag{105}$$

where

$$\tilde{\psi}_e(\tilde{y}_e) = \sqrt{2\pi s^2}\exp\big(-\frac{s^2}{2}\tilde{y}_e^2\big) \tag{106}$$

$$\propto \exp\big(-\frac{s^2}{2}\tilde{y}_e^2\big) \tag{107}$$

and

$$\tilde{\phi}_v(\tilde{x}_v) = \exp\big(-\frac{\sigma^2}{2}\tilde{x}_v^2\big) \tag{108}$$

are the Fourier transforms of (101) and (102) respectively.[3]

The corresponding dual NFG is shown in Fig. 2, where factors (107) are represented by the big unlabeled boxes and factors (108) are represented by the small unlabeled boxes.

A straightforward generalization of Proposition 1 gives the following local mapping that relates the edge marginal probabilities of the primal NFG to the edge marginal probabilities in the dual NFG

$$\frac{f_{\mathrm{p},e}(y_e)}{\psi_e(y_e)} = \Big(\mathcal{F}\frac{f_{\mathrm{d},e}}{(\mathcal{F}\psi_e)}\Big)(\tilde{y}_e) \tag{109}$$

$$= \Big(\mathcal{F}\frac{f_{\mathrm{d},e}}{\tilde{\psi}_e}\Big)(\tilde{y}_e) \tag{110}$$

In other words, the functions $f_{\mathrm{p},e}(t)/\psi_e(t), t\in\mathbb{R}$ and $f_{\mathrm{d},e}(\tilde{t})/\tilde{\psi}_e(\tilde{t}), \tilde{t}\in\mathbb{R}$ are Fourier pairs. Similarly, it holds that

$$\frac{f_{\mathrm{p},v}(x_v)}{\phi_v(x_v)} = \Big(\mathcal{F}\frac{f_{\mathrm{d},v}}{\tilde{\phi}_v}\Big)(\tilde{x}_v) \tag{111}$$

as a generalization of Proposition 2.

Again, as a consequence of (110) and (111), it is possible to estimate the edge/vertex marginal densities in the dual domain, and then transform them to the primal domain. Let

3. Following (104), the Fourier transform of $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\big(-\frac{1}{2\sigma^2}x^2\big)$, i.e., the PDF of a Gaussian distribution with mean zero and with standard deviation σ, is given by $(\mathcal{F}f)(\tilde{x}) = \exp\big(-\frac{\sigma^2}{2}\tilde{x}^2\big)$. However, the scale factors that appear in the Fourier transforms of (101) and (102) are unimportant in our formulation. After all, we are mainly concerned with computing the marginal densities of (100) from the corresponding marginal densities of (105).

us denote the estimated marginal variance in the dual domain by $\hat{\sigma}^2_{\mathrm{d}}$, which can be mapped to the primal NFG via (111). Indeed

$$f_{\mathrm{p},v}(x_v) = \phi_v(x_v)\Big(\mathcal{F}\frac{f_{\mathrm{d},v}}{\tilde{\phi}_v}\Big)(\tilde{x}_v) \tag{112}$$

$$\propto \exp(-\frac{1}{2\sigma^2}x_v^2)\int_{-\infty}^{\infty}\frac{\exp(-\frac{1}{2\hat{\sigma}^2_{\mathrm{d}}}\tilde{x}_v^2)}{\exp(-\frac{\sigma^2}{2}\tilde{x}_v^2)})\mathrm{e}^{-\mathrm{i}\,\tilde{x}_v x_v}d\tilde{x}_v \tag{113}$$

$$\propto \exp(-\frac{1}{2\sigma^2(1-\sigma^2\hat{\sigma}^2_{\mathrm{d}})}x_v^2) \tag{114}$$

Let $f_{\mathrm{p},v}(x_v) \propto \exp(-\frac{1}{2\hat{\sigma}^2_{\mathrm{d}\to\mathrm{p}}}x_v^2)$, where $\hat{\sigma}^2_{\mathrm{d}\to\mathrm{p}}$ is the estimated marginal variance in the primal domain. After matching the exponents on the left-hand side and the-right hand side of (114), we get

$$\hat{\sigma}^2_{\mathrm{d}\to\mathrm{p}} = \sigma^2(1-\sigma^2\hat{\sigma}^2_{\mathrm{d}}), \tag{115}$$

which can be used to compute an estimate of the marginal variance in the primal domain from the estimated marginal variance in the dual domain.

9.1 Numerical Experiments

We consider an $N = 15\times 15$ Gaussian Markov random field with periodic boundary conditions and with PDF given by (100). In both domains (i.e., the primal and the dual domains), we employ the Gibbs sampling algorithm with a systematic (i.e., deterministic) sweep strategy (Robert and Casella, 2004, Chapter 10) to estimate the marginal densities.[4]

We set $\sigma = 5$ and the number of samples $L = 10^3$ in all the experiments. The exact value of σ^2_{p} is computed by inverting the information (i.e., precision) matrix associated with the model, which is feasible for this size of graph. Notice that the total number of variables is $|\mathcal{V}| = N = 225$ in the primal domain, and is $|\mathcal{E}| = 2N = 450$ in the dual domain.

In Fig. (14), the dotted black paths show $\hat{\sigma}^2_{\mathrm{p}}$ the estimated marginal variance obtained directly in the primal NFG and the dashed blue paths show $\hat{\sigma}^2_{\mathrm{d}\to\mathrm{p}}$ the estimated marginal variance in the primal NFG obtained from $\hat{\sigma}^2_{\mathrm{d}}$, and then mapped to the primal domain via (115). In all plots the horizontal axis shows the number of samples. In order to compare better the convergence, the right panel of Fig. (14) illustrates the zoomed plots of the left panel in the range $[10^2, 10^3]$.

In Fig. (14)-Top, we set the intervariable variance $s = 1$. The exact value of σ^2_{p} is about 0.5589. The Gibbs sampling algorithm converges quickly in the primal NFG, where after 10^2 samples the relative error is about 10^{-1}. However, the Gibbs sampler suffers from slow and erratic convergence in the dual NFG. We set $s = 20$ in Fig. (14)-Middle. In this example σ^2_{p} is about 20.2046. We observe that increasing s improves the convergence in the dual domain, but degrades the convergence in the primal domain. In the last experiment, we set

4. Applying the Gibbs sampling algorithm to Gaussian Markov random fields is straightforward and can be represented in matrix form. The algorithm can be viewed as the stochastic generalization of the Gauss-Seidel algorithm, as pointed out in (Goodman and Sokal, 1989, Section VIII).

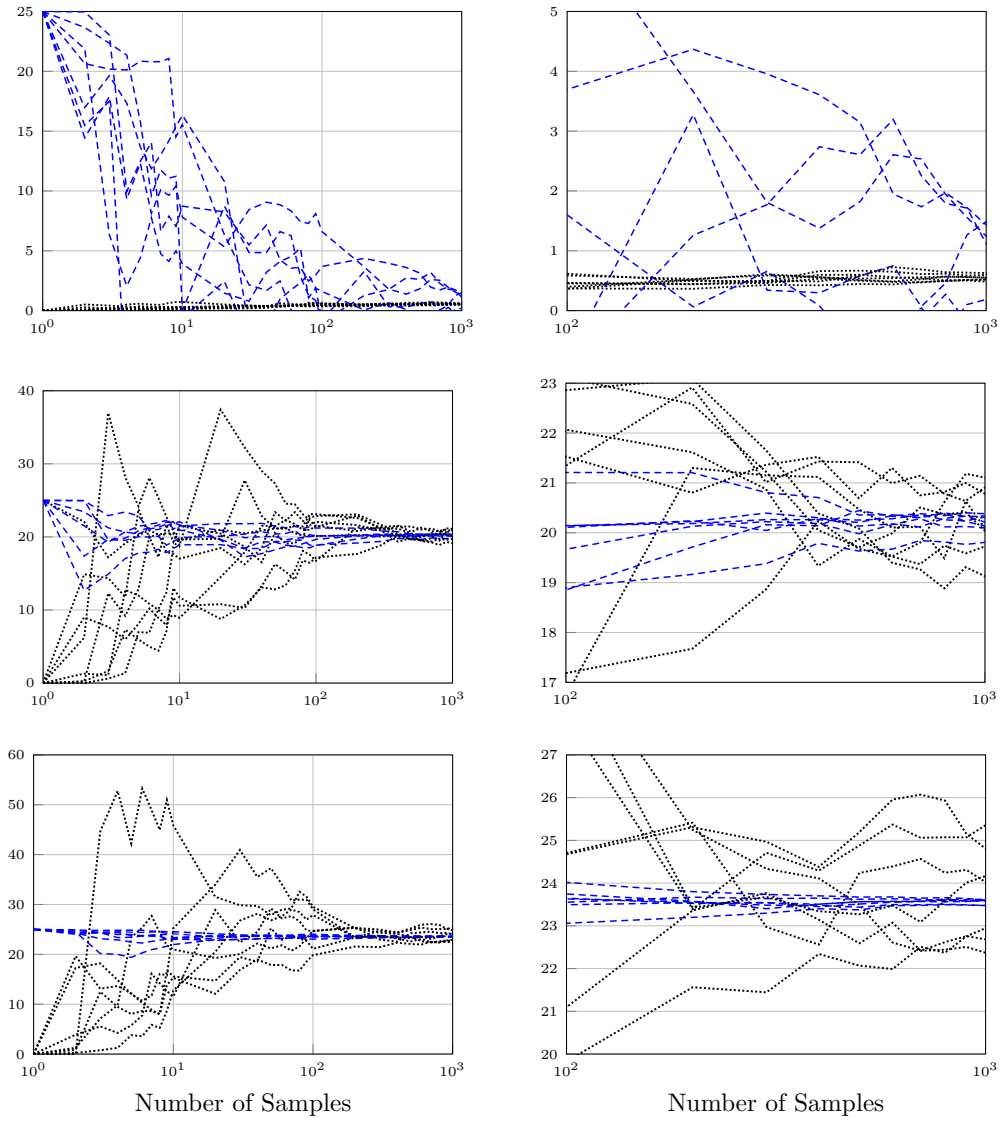

Figure 14: Estimated marginal variances as a function of the number of samples L for a 15×15 Gaussian Markov random field with PDF as in (100), with periodic boundaries, and with $\sigma = 5$. The right panel illustrates the zoomed plots of the left panel for $L \in [10^2, 10^3]$. The dotted black paths and the dashed blue paths show $\hat{\sigma}^2_{\mathrm{p}}$ and $\hat{\sigma}^2_{\mathrm{d}\to\mathrm{p}}$ respectively. The intervariable variance is set to: (Top) $s = 1$, (Middle) $s = 20$, and (Bottom) $s = 40$.

$s = 40$. In this case, the exact value of σ^2_{p} is about 23.5498. Simulation results are shown in Fig. (14)-Bottom, where, in contrast to Fig. (14)-Top, convergence in the primal domain is slow. However, the Gibbs sampling algorithm in the dual domain can provide very accurate

estimates of the marginal variance. A relative error of about 1.5×10^{-3} is achieved using only 10^2 samples.

Intuitively, as s^2 grows, the generated samples in the primal domain may vary widely, which leads to unstable estimates of σ_{p}^2. The situation is the opposite in the dual domain, as in the Fourier transform of (101) the standard deviation s is replaced by $1/s$, cf. (107). The potential advantages of performing the computations in the dual domain needs to be further investigated for other choices of parameters, e.g., by fixing s^2 and varying the vertex variance σ^2.

10. Conclusion

We proved that the edge/vertex marginals densities of a primal NFG and the corresponding edge/vertex marginal densities of its dual NFG are related via local mappings. The mapping provides a simple procedure to transform simultaneously the estimated marginals from one domain to the other. Furthermore, the mapping relies on no assumptions on the size or on the topology of the graphical model. Details of the mapping, including its fixed points, were derived for the Ising model and were extended to non-binary models (e.g., the Potts model) and to continuous models. We discussed that the subgraphs-world process can be employed as a rapidly mixing Markov chain to generate configurations in the dual NFG of the ferromagnetic Ising models in a positive external field. In this case, the edge marginals densities of the dual NFG can first be obtained in polynomial time in the dual domain, and then be transformed to the primal domain using the proposed mappings.

Our numerical experiments show that estimating the marginals densities in the dual NFG can sometimes significantly improve the quality of approximations in terms of relative error. E.g., variational algorithms in the dual NFG of the ferromagnetic Ising and Potts models converge faster in the low-temperature regime (i.e., when the coupling parameters are large), and, for some settings, the Gibbs sampling algorithm in the dual domain can give more accurate estimates of the marginal densities of Gaussian Markov random fields with the thin-membrane prior.

Following Remark 4, the factors in the dual NFG of the Ising model can in general take negative values. In principle, marginal functions can still be estimated via the BP algorithm in these cases. However, we observed that in many such examples (e.g., spin glasses and antiferromagnetic Ising models) BP suffers from slow and erratic convergence in the dual domain. Analyzing the dynamics of the BP algorithm and designing new and improved variational algorithms in NFGs with negative (or complex) factors is left for future work.

Appendix A: Factor Graphs and normal relizations

A factor graph is a diagram that represents the factorization of a multivariate function, in which there is a variable node for each variable and a factor node for each factor. An edge connects the variable node of x_i to the factor node of $\psi_j(\cdot)$ if and only if x_i is an argument of $\psi_j(\cdot)$. See (Kschischang et al., 2001), for more details.

As an example, let $\mathcal{A} = \mathbb{Z}/2\mathbb{Z}$ and assume that $\pi_{\mathrm{p}}(\cdot)$ can be factored as

$$\pi_{\mathrm{p}}(x_1, x_2, x_3) \propto \phi_1(x_1)\phi_2(x_2)\phi_3(x_3)\psi_1(x_1, x_2)\psi_1(x_2, x_3)\psi_1(x_1, x_3) \tag{116}$$

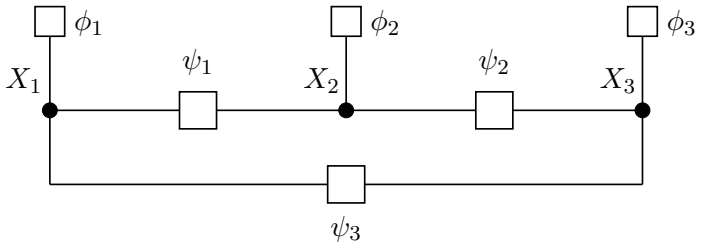

Figure 15: The factor graph of (116). The filled circles show the variable nodes and the empty boxes represent the factor nodes.

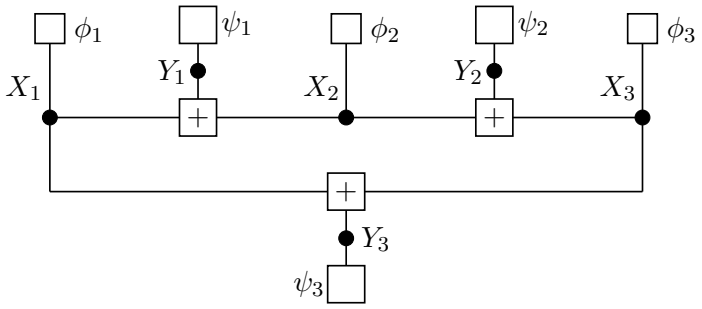

Figure 16: The factor graph of (117). The filled circles show the variable nodes and the empty boxes represent the factor nodes. The boxes labeled "+" are instances of zero-sum indicator factors given by (2).

up to scale. The factor graph in Fig. (15) expresses the factorization in (116), where the filled circles show the variable nodes and the empty boxes represent the factor nodes.

If we further assumed that $\psi_1(\cdot)$, $\psi_2(\cdot)$, and $\psi_3(\cdot)$ are only functions of the difference between their two arguments, we can express $\pi_{\mathrm{p}}(\cdot)$ by the following factorization

$$\pi_{\mathrm{p}}(x_1, x_2, x_3) \propto \phi_1(x_1)\phi_2(x_2)\phi_3(x_3)\psi_1(y_1)\psi_1(y_2)\psi_1(y_3), \tag{117}$$

where $y_1 = x_1 - x_2$, $y_2 = x_2 - x_3$, and $y_3 = x_3 - x_1$.

The factor graph that corresponds to (117) is shown in Fig. (16). The boxes labeled "+" are instances of zero-sum indicator factors given by (2), which impose the constraint that all their incident variables sum to zero modulo two.

By contrast, in NFGs, introduced in (Forney, 2001), variable are represented by edges. The NFG that corresponds to (117) is illustrated in Fig. (17). Since variables y_1, y_2, and y_3 appear in only two factors, they are easily represented by their corresponding edges. However, variables x_1, x_2, and x_3 that appear in more than two factors are replicated via equality indicator factors. E.g., in Fig. (17)

$$\delta_{=}(x_1, x_1', x_1'') = \begin{cases} 1, & \text{if } x_1 = x_1' = x_1'' \\ 0, & \text{otherwise,} \end{cases} \tag{118}$$

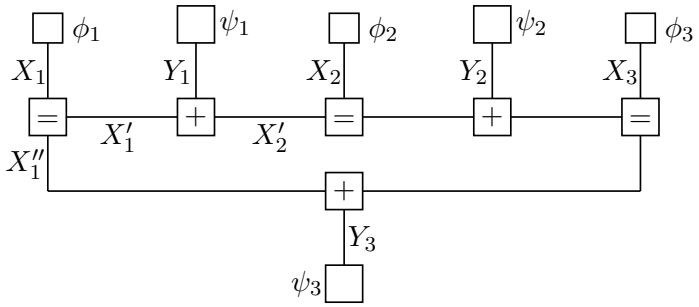

Figure 17: The NFG of (117). The unlabeled boxes represent the factors $\phi(\cdot)$ and $\psi(\cdot)$. Boxes labeled "=" are instances of equality indicator factors given by (1), and the boxes labeled "+" are instances of zero-sum indicator factors as in (2)

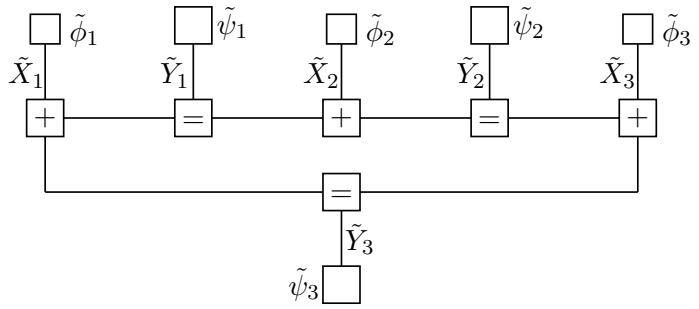

Figure 18: The dual of the NFG in Fig. 17.

which can also be expressed as $\delta_=(x_1, x_1', x_1'') = \delta(x_1 - x_1')\delta(x_1 - x_1'')$. Here $\delta(\cdot)$ denotes either the Kronecker delta function (in discrete models discussed in Sections 3 and 4) or the Dirac delta function (in continuous models discussed in Section 9). Replicating a variable by an equality indicator factor allows us to enforce the condition that each variable appears in (at most) two factors. As a result, in NFGs there is an edge for every variable (Loeliger, 2004; Forney, 2011).

The dual NFG can be obtained by replacing each variable by its corresponding dual variable and each factor by its 1D DFT, cf. Section 4. The dual of the NFG in Fig. 17 is illustrated in Fig. 18.

Appendix B: Marginal densities of the 1D Ising and Potts models

The 1D Ising and Potts models with free or with periodic boundary conditions are exactly solved models. It is therefore possible to compute the edge marginal probabilities exactly in these models via the sum-product algorithm (Kschischang et al., 2001) or via the transfer matrix method (Baxter, 2007). Here we employ our proposed mappings as an alternative (and somewhat simpler) way to compute such marginals.

The Primal NFG of the 1D Ising model in zero field and with periodic boundaries is shown in Fig. 19, where the unlabeled boxes represent (7). To construct the dual NFG of the model, variables are replaced by their corresponding dual variable, factors (7) are replaced

by (23), equality indicator factors are replaced by zero-sum indicator factors and vice-versa. The "$\circ$" symbols are immaterial as $\mathcal{A} = \mathbb{Z}/2\mathbb{Z}$. Fig. 20 shows the dual of Fig. 19.

In the dual NFG $\tilde{X}_1 = \tilde{X}_2 = \ldots = \tilde{X}_{|\mathcal{E}|}$, therefore there are only two valid configurations, namely the all-zeros and the all-ones configurations. Hence

$$Z_{\mathrm{d}} = 2^{|\mathcal{E}|}\big(\prod_{e\in\mathcal{E}} \cosh(\beta J_e) + \prod_{e\in\mathcal{E}} \sinh(\beta J_e)\big), \tag{119}$$

which gives

$$\pi_{\mathrm{d},e}(0) = \frac{\prod_{e\in\mathcal{E}} \cosh(\beta J_e)}{\prod_{e\in\mathcal{E}} \cosh(\beta J_e) + \prod_{e\in\mathcal{E}} \sinh(\beta J_e)} \tag{120}$$

$$= \frac{1}{1 + \prod_{e\in\mathcal{E}} \tanh(\beta J_e)} \tag{121}$$

and

$$\pi_{\mathrm{d},e}(1) = \frac{\prod_{e\in\mathcal{E}} \tanh(\beta J_e)}{1 + \prod_{e\in\mathcal{E}} \tanh(\beta J_e)} \tag{122}$$

From (49), we obtain the edge marginal probabilities of the primal NFG as

$$\pi_{\mathrm{p},e}(0) = \frac{\mathrm{e}^{\beta J_e}}{2\cosh(\beta J_e)} \frac{1 + \prod_{e'\in\mathcal{E}\setminus\{e\}} \tanh(\beta J_{e'})}{1 + \prod_{e\in\mathcal{E}} \tanh(\beta J_e)} \tag{123}$$

$$\pi_{\mathrm{p},e}(1) = \frac{\mathrm{e}^{-\beta J_e}}{2\cosh(\beta J_e)} \frac{1 - \prod_{e'\in\mathcal{E}\setminus\{e\}} \tanh(\beta J_{e'})}{1 + \prod_{e\in\mathcal{E}} \tanh(\beta J_e)} \tag{124}$$

In the dual NFG of the 1D Ising model with free boundary conditions, there is only one valid configuration, namely the all-zeros configuration. Consequently

$$\pi_{\mathrm{d},e}(0) = 1 - \pi_{\mathrm{d},e}(1) = 1 \tag{125}$$

and therefore from (49) we obtain

$$\pi_{\mathrm{p},e}(0) = \frac{\mathrm{e}^{\beta J_e}}{2\cosh(\beta J_e)} \tag{126}$$

$$\pi_{\mathrm{p},e}(1) = \frac{\mathrm{e}^{-\beta J_e}}{2\cosh(\beta J_e)} \tag{127}$$

for $e \in \mathcal{E}$. Due to symmetry

$$\pi_{\mathrm{p},v}(0) = \frac{1}{2}\pi_{\mathrm{p},e}(0) + \frac{1}{2}\pi_{\mathrm{p},e}(1) \tag{128}$$

for $v \in \mathcal{V}$. Thus

$$\pi_{\mathrm{p},v}(0) = \pi_{\mathrm{p},v}(1) = \frac{1}{2} \tag{129}$$

The marginal probability $\pi_{\mathrm{p},e}(0)$ in (126) attains the lower bound in (51). We also observe that in the thermodynamic limit our results are independent of the boundary conditions. In other words, the effect of the boundary conditions becomes negligible as $|\mathcal{V}|, |\mathcal{E}| \to \infty$.

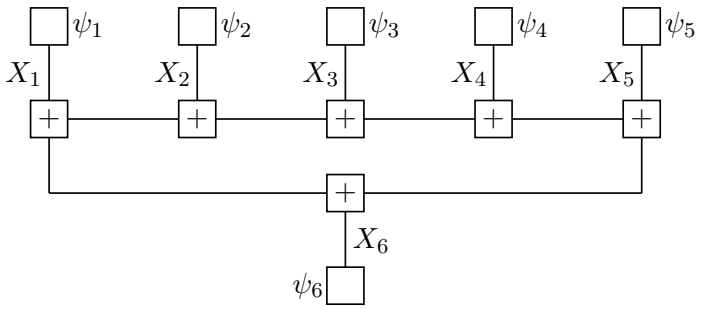

Figure 19: The primal NFG of the 1D Ising model in the absence of an external field and with periodic boundary conditions. The unlabeled boxes represent (7), boxes labeled "=" are instances of equality indicator factors as in (1), the boxes labeled "+" are instances of zero-sum indicator factors given by (2).

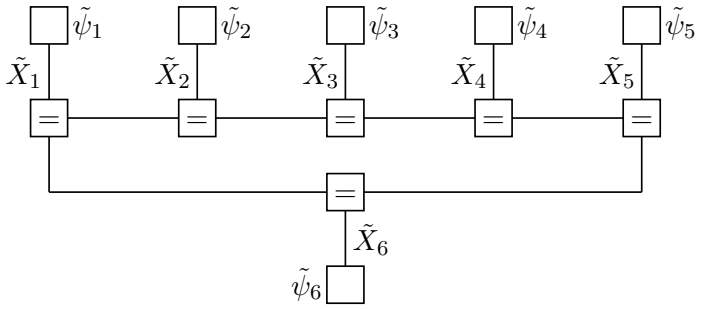

Figure 20: The dual of the NFG in Fig. 19, where the unlabeled boxes represent (23).

Finally, we briefly discuss similar mappings for the edge marginal probabilities of the 1D homogeneous q-state Potts model with periodic boundaries.

In the dual NFG of the model, the unlabeled boxes represent (17). There are exactly q configurations in the dual domain, each with $\tilde{X}_1 = \tilde{X}_2 = \ldots = \tilde{X}_{|\mathcal{E}|}$. Hence

$$Z_{\mathrm{d}} = (\mathrm{e}^{\beta J} + q - 1)^{|\mathcal{E}|} + (q-1)(\mathrm{e}^{\beta J} - 1)^{|\mathcal{E}|}, \tag{130}$$

where the first term is the contribution of the all-zeros configuration, and the contribution of each of the remaining $q-1$ configurations to the partition function is $(\mathrm{e}^{\beta J} - 1)^{|\mathcal{E}|}$.

Hence

$$\pi_{\mathrm{d},e}(0) = \frac{(\mathrm{e}^{\beta J} + q - 1)^{|\mathcal{E}|}}{(\mathrm{e}^{\beta J} + q - 1)^{|\mathcal{E}|} + (q-1)(\mathrm{e}^{\beta J} - 1)^{|\mathcal{E}|}} \tag{131}$$

and for $t \in \{1, 2, \ldots, q-1\}$

$$\pi_{\mathrm{d},e}(t) = \frac{(\mathrm{e}^{\beta J} - 1)^{|\mathcal{E}|}}{(\mathrm{e}^{\beta J} + q - 1)^{|\mathcal{E}|} + (q-1)(\mathrm{e}^{\beta J} - 1)^{|\mathcal{E}|}} \tag{132}$$

From (77) we obtain

$$\pi_{\mathrm{p},e}(0) = \mathrm{e}^{\beta J} \frac{(\mathrm{e}^{\beta J}+q-1)^{|\mathcal{E}|-1} + (q-1)(\mathrm{e}^{\beta J}-1)^{|\mathcal{E}|-1}}{(\mathrm{e}^{\beta J}+q-1)^{|\mathcal{E}|} + (q-1)(\mathrm{e}^{\beta J}-1)^{|\mathcal{E}|}} \tag{133}$$

The other entries of $(\pi_{\mathrm{p},e}(a), a \in \mathcal{A})$ can be computed analogously.

Acknowledgments

The author is extremely grateful to G. David Forney, Jr., for his comments and for his continued support. The author wishes to thank Justin Dauwels for his comments on an earlier draft of this paper. The author would like to thanks the editor, Mohammad Emtiyaz Khan, and the anonymous reviewers for their helpful feedback.

References

A. Al-Bashabsheh and Y. Mao. Normal factor graphs and holographic transformations. *IEEE Transactions on Information Theory*, 57:752–763, 2011.

A. Al-Bashabsheh and Y. Mao. On stochastic estimation of the partition function. *Proc. IEEE International Symposium on Information Theory*, pages 1504–1508, 2014.

R. J. Baxter. *Exactly Solved Models in Statistical Mechanics*. Dover Publications, 2007.

J. Besag and P. J. Green. Spatial statistics and Bayesian computation. *Journal of the Royal Statistical Society: Series B*, 55:25–37, 1993.

R. N. Bracewell. *The Fourier Transform and Its Applications*. McGraw-Hill, 1999.

J.-Y. Cai and X. Chen. *Complexity Dichotomies for Counting Problems: Volume 1, Boolean Domain*. Cambridge University Press, 2017.

J.-Y. Cai, P. Lu, and M. Xia. Holographic algorithms by Fibonacci gates and holographic reductions for hardness. *Proc. IEEE Symposium on Foundations of Computer Science*, pages 644–653, 2008.

G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová. Machine learning and the physical sciences. *Reviews of Modern Physics*, 91:045002, 2019.

G. Cossu, L. Del Debbio, T. Giani, A. Khamseh, and M. Wilson. Machine learning determination of dynamical parameters: The Ising model case. *Physical Review B*, 100: 064304, 2019.

P. Dagum and M. Luby. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence*, 60:141–153, 1993.

G. D. Forney. Codes on graphs: Normal realization. *IEEE Transactions on Information Theory*, 47:520–548, 2001.

G. D. Forney. Codes on graphs: Duality and MacWilliams identities. *IEEE Transactions on Information Theory*, 57:1382–1397, 2011.

G. D. Forney. Codes on graphs: Models for elementary algebraic topology and statistical physics. *IEEE Transactions on Information Theory*, 64:7465–7487, 2018.

J. Goodman and A. D. Sokal. Multigrid Monte Carlo method. Conceptual foundations. *Physical Review D*, 40:2035–2070, 1989.

P. Horowitz and W. Hill. *The Art of Electronics*. Cambridge University Press, 1989.

K. Huang. *Statistical Mechanics*. John Wiley & Sons, 1987.

M. Jerrum and A. Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on Computing*, 22:1087–1116, 1993.

T. Kihara, Y. Midzuno, and T. Shizume. Statistics of two-dimensional lattices with many components. *Journal of the Physical Society of Japan*, 9:681–687, 1954.

J. M. Kosterlitz and D. J. Thouless. Ordering, metastability and phase transitions in two-dimensional systems. *Journal of Physics C: Solid State Physics*, 6:1181–1203, 1973.

H. A. Kramers and G. H. Wannier. Statistics of the two-dimensional ferromagnet. Part I. *Physical Review*, 60:252–262, 1941.

F. R. Kschischang, B. J. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47:498–519, 2001.

S. Kudekar and N. Macris. Decay of correlations for sparse graph error correcting codes. *SIAM Journal on Discrete Mathematics*, 25:956–988, 2011.

J. Liu, X. Wang, and X.-C. Tai. Deep convolutional neural networks with spatial regularization, volume and star-shape priors for image segmentation. *Journal of Mathematical Imaging and Vision*, 2022. doi: 10.1007/s10851-022-01087-x.

H.-A. Loeliger. An introduction to factor graphs. *IEEE Signal Processing Magazine*, 29: 28–41, 2004.

D. M. Malioutov. *Approximate Inference in Gaussian Graphical Models*. PhD thesis, Massachusetts Institute of Technology, 2008.

F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. *Communications in Mathematical Physics*, 161:447–486, 1994.

C. A. McGrory, D. M. Titterington, R. Reeves, and A. N. Pettitt. Variational Bayes for estimating the parameters of a hidden Potts model. *Statistics and Computing*, 19:329–340, 2009.

M. Molkaraie. The primal versus the dual Ising model. *Proc. 55th Annual Allerton Conference on Communication, Control, and Computing*, pages 53–60, 2017.

M. Molkaraie. Marginal densities, factor graph duality, and high-temperature series expansions. *Proc. International Conference on Artificial Intelligence and Statistics*, pages 256–265, 2020.

M. Molkaraie and V. Gómez. Monte Carlo methods for the ferromagnetic Potts model using factor graph duality. *IEEE Transactions on Information Theory*, 59:7449–7464, 2018.

M. Molkaraie and H.-A. Loeliger. Partition function of the Ising model via factor graph duality. *Proc. IEEE International Symposium on Information Theory*, pages 2304–2308, 2013.

J. M. Mooij. libdai: A free and open source C++ library for discrete approximate inference in graphical models. *The Journal of Machine Learning Research*, 11:2169–2173, 2010.

A. Morningstar and R. G. Melko. Deep learning the Ising model near criticality. *The Journal of Machine Learning Research*, 18:1–17, 2018.

K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

G. F. Newell and E. W. Montroll. On the theory of the Ising model of ferromagnetism. *Reviews of Modern Physics*, 25:353–389, 1953.

H. Nishimori and G. Ortiz. *Elements of Phase Transitions and Critical Phenomena*. Oxford University Press, 2015.

L. Onsager. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Physical Review*, 65:117–149, 1944.

R. B. Potts. Some generalized order-disorder transformations. *Proc. the Cambridge Philosophical Society*, 48:106–109, 1952.

T. Richardson and R. Urbanke. *Modern Coding Theory*. Cambridge University Press, 2008.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2004.

A. Terras. *Fourier Analysis on Finite Groups and Applications*. Cambridge University Press, 1999.

L. G. Valiant. Holographic algorithms. *SIAM Journal on Computing*, 37:1565–1594, 2008.

M. J. Wainwright and M. I. Jordan. *Graphical Models, Exponential Families, and Variational Inference*. Now Publishers Inc, 2008.

Y. Weiss and W. T. Freeman. Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Advances in Neural Information Processing Systems*, pages 673–679, 2000.

G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*. Springer, 1995.

F. Y. Wu. The Potts model. *Reviews of Modern Physics*, 54:235–268, 1982.

J. M. Yeomans. *Statistical Mechanics of Phase Transitions*. Oxford University Press, 1992.

H. Yu, S. Wu, and J. Dauwels. Efficient variational Bayes learning of graphical models with smooth structural changes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. doi: 10.1109/TPAMI.2022.3140886.