Modified Bethe Permanent of a Nonnegative Matrix

Shashank Vatedka*, Pascal Vontobel†
*Dept. of Electrical Engineering, Indian Institute of Technology Hyderabad, India
†Dept. of Information Engineering, The Chinese University of Hong Kong, Hong Kong SAR

***Abstract*—Currently the best deterministic polynomial-time algorithm for approximating the permanent of a non-negative matrix is based on minimizing the Bethe free energy function of a certain normal factor graph (NFG). In order to improve the approximation guarantee, we propose a modified NFG with fewer cycles, but still manageable function-node complexity; we call the approximation obtained by minimizing the function of the modified normal factor graph the modified Bethe permanent. For nonnegative matrices of size 3×3, we give a tight characterization of the modified Bethe permanent. For non-negative matrices of size $n \times n$ with $n \geqslant 3$, we present a partial characterization, along with promising numerical results. The analysis of the modified NFG is also interesting because of its tight connection to an NFG that is used for approximating a permanent-like quantity in quantum information processing.**

I. Introduction

Given a matrix $\Theta = (\theta_{ij})_{i,j}$ of size $n \times n$, the permanent of Θ is defined to be

$$\mathrm{perm}(\Theta) := \sum_{\sigma \in S_n} \prod_{i=1}^{n} \theta_{i\sigma(i)}, \tag{1}$$

where S_n denotes the symmetric group on n elements. In this paper, we will assume that Θ is a nonnegative matrix (i.e., $\theta_{ij} \geqslant 0$ for all i, j) and that $\mathrm{perm}(\Theta) > 0$.

Computing the permanent has several applications including finding the number of perfect matchings in a bipartite graph, particle tracking in images [1], and maximum-likelihood estimation of histograms of distributions [2]. Unfortunately, the best known deterministic algorithm [3] for $\mathrm{perm}(\Theta)$ takes time $O(n2^n)$ for the worst-case instance even if we restrict the entries of Θ to be binary, and computing the permanent of a binary matrix is known to be #P-complete [4]. Several approximation algorithms have been proposed [5], [6], [7], and the best known randomized algorithm [8] uses Markov-Chain Monte-Carlo methods. The best known deterministic approximation algorithms are sum-product algorithm (SPA) based methods. See [1], [9], [10] and references therein.

In the language of [10], the papers [1], [9], [10] formulated a normal factor graph (NFG) $\mathsf{N}_{\mathrm{B}}(\Theta)$ whose underlying topology is that of a complete bipartite graph with two times n vertices and whose partition sum equals $\mathrm{perm}(\Theta)$ for a given matrix Θ. The approximation of $\mathrm{perm}(\Theta)$ that is obtained by minimizing the Bethe free energy function [11] associated with $\mathsf{N}_{\mathrm{B}}(\Theta)$ is called the Bethe permanent and denoted by $\mathrm{perm}_{\mathrm{B}}(\Theta)$. For stating results about the Bethe permanent, it turns out to be useful to define $\eta_{\mathrm{B}}(\Theta) := \frac{1}{n} \log \frac{\mathrm{perm}(\Theta)}{\mathrm{perm}_{\mathrm{B}}(\Theta)}$, where Θ is a nonnegative matrix of size $n \times n$, and the logarithm is taken to the base two.

It was shown in [10] that the Bethe free energy function [11] and the SPA are very well behaved for the NFG $\mathsf{N}_{\mathrm{B}}(\Theta)$. In particular, the Bethe free energy function is, when suitably parameterized, a convex function. Moreover, the SPA converges to the minimum of the Bethe free energy function. Finally, a series of papers established that $0 \leqslant \eta_{\mathrm{B}}(\Theta) \leqslant 1/2$, where both the lower and the upper bounds are tight. Namely, Gurvits [12] showed that $\eta_{\mathrm{B}}(\Theta) \geqslant 0$ for all Θ. (An alternative proof of this was recently given in [13].) Moreover [12] conjectured that $\eta_{\mathrm{B}}(\Theta) \leqslant 1/2$. Later, [14] showed that $\eta_{\mathrm{B}}(\Theta) \leqslant 1$. Very recently, [15] showed that, indeed, $\eta_{\mathrm{B}}(\Theta) \leqslant 1/2$. When Gurvits [12] conjectured that $\eta_{\mathrm{B}}(\Theta) \leqslant 1/2$, he noticed that this bound holds with equality for the Kronecker product of the $\frac{n}{2} \times \frac{n}{2}$ identity matrix with the 2×2 all-ones matrix $1_{2\times 2}$.

In this paper, we propose a modified NFG $\mathsf{N}_{\mathrm{MB}}(\Theta)$ that overcomes this worst-case instance. ("MB" stands for "modified Bethe".) The resulting Bethe approximation will be called $\mathrm{perm}_{\mathrm{MB}}(\Theta)$ and its approximation of the true permanent will be characterized in terms of $\eta_{\mathrm{MB}}(\Theta) := \frac{1}{n} \log \frac{\mathrm{perm}(\Theta)}{\mathrm{perm}_{\mathrm{MB}}(\Theta)}$. The modified NFG $\mathsf{N}_{\mathrm{MB}}(\Theta)$ strikes a good balance between removing cycles in $\mathsf{N}_{\mathrm{B}}(\Theta)$ and not increasing the function node complexity by too much.[1]

We will study the NFG $\mathsf{N}_{\mathrm{MB}}(\Theta)$, in particular the associated Bethe free energy function and the behavior of the SPA when operating on $\mathsf{N}_{\mathrm{MB}}(\Theta)$.

- For $n = 2$, the NFG $\mathsf{N}_{\mathrm{MB}}(\Theta)$ has no cycles and therefore $\eta_{\mathrm{MB}}(\Theta) = 0$. This case will not be discussed further.
- For $n = 3$, we show that the NFG $\mathsf{N}_{\mathrm{MB}}(\Theta)$ is very well behaved in the sense that the Bethe free energy function is convex, that the SPA converges to the minimum of the Bethe free energy function, and that $0 \leqslant \eta_{\mathrm{MB}}(\Theta) \leqslant 1/3$, where both the lower and the upper bounds are tight.
- For $n > 3$, empirical results show that $\eta_{\mathrm{MB}}(\Theta)$ is always much less than $1/2$. Moreover, the SPA converges to an

This work was done when the first author was a postdoc at the Dept. of Information Engineering, The Chinese University of Hong Kong, where he was supported in part by CUHK Direct Grants 4055039 and 4055077. The work described in this paper was partially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Nos. CUHK 14209317 and CUHK 14207518).

978-1-7281-8895-9/20/$31.00 © 2020 IEEE

[1] Our approach at obtaining a better approximation can be seen as somewhat similar to region approximations [11]. For such region approximations it is clear that the larger the regions are, the better the approximation will be. However, the complexity will also grow exponentially with the region sizes.

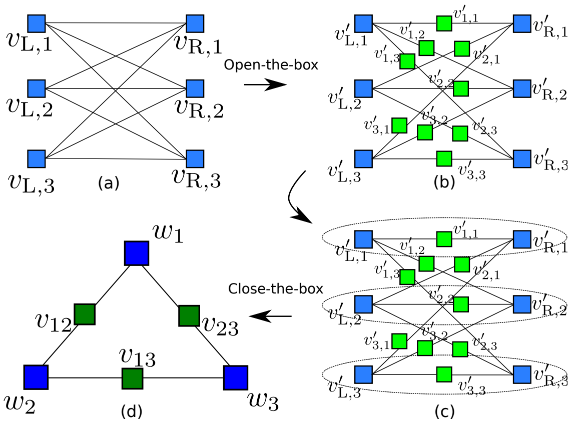

Fig. 1: From $\mathsf{N}_{\mathrm{B}}(\Theta)$ to $\mathsf{N}_{\mathrm{MB}}(\Theta)$ for $n = 3$. (See text for details.)

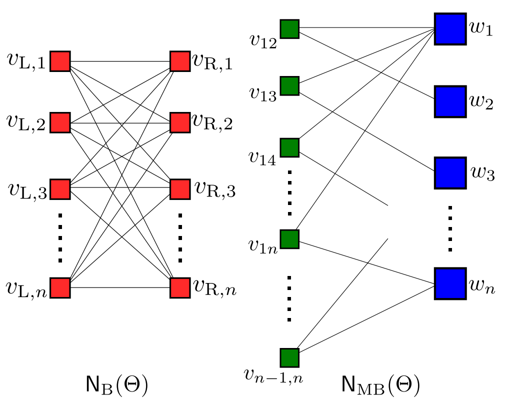

Fig. 2: $\mathsf{N}_{\mathrm{B}}(\Theta)$ and $\mathsf{N}_{\mathrm{MB}}(\Theta)$ for general n.

interior point of the local marginal polytope when none of the vertices are optimal. This is despite the fact that we can establish that the Bethe free energy function is not convex everywhere and despite the fact that there are vertices of the local marginal polytope where the directional derivative has infinite slope. This means that vertices with directional derivatives with infinite slope have also directional derivatives with finite slope, thereby allowing the SPA to escape a non-optimal vertex. (This behavior is in contrast to binary LDPC codes with bit node degree at least three, whose Bethe free energy function is such that at codeword vertices of the local marginal polytope all directional derivatives have infinite slope.)

Let us highlight two aspects which make $\mathsf{N}_{\mathrm{MB}}(\Theta)$ a worthwhile object of study:

- Because $\mathsf{N}_{\mathrm{MB}}(\Theta)$ breaks some symmetry in $\mathsf{N}_{\mathrm{B}}(\Theta)$, it is advisable to reorder the rows and columns of Θ (which leaves the permanent invariant) towards optimizing the properties of $\mathsf{N}_{\mathrm{B}}(\Theta)$. In particular, it is advisable to permute the rows and columns of Θ such that $\prod_{i=1}^{n} \theta_{i\sigma(i)}$ is maximized by the trivial permutation σ in S_n. The required row and column permutations can be found with the help of an algorithm that finds the maximum-weight matching. This can be accomplished efficiently, e.g., by running the max-product or min-sum algorithm on a suitable NFG (see [16], [17] for details). Such a low-complexity pre-processing of an NFG might also be interesting in other contexts.
- There is a tight connection between $\mathsf{N}_{\mathrm{MB}}(\Theta)$ and an NFG in [18] that was used for approximating a permanent-like quantity in quantum information processing. Similar NFGs have been used to represent quantum systems with multiple measurements [19], [20] and estimate rates of quantum channels with memory [21]. Insights gained for $\mathsf{N}_{\mathrm{MB}}(\Theta)$ are also relevant for these other NFGs.

The rest of the paper is structured as follows. In Section II we introduce $\mathsf{N}_{\mathrm{MB}}(\Theta)$. In Section III we discuss an efficient implementation of the SPA for $\mathsf{N}_{\mathrm{MB}}(\Theta)$. In Section IV we analyze $\mathsf{N}_{\mathrm{MB}}(\Theta)$ for $n = 3$ and in Section V for $n > 3$. Finally, in Section VI we discuss some simulation results. Due to a lack of space, we have omitted proofs and detailed calculations, which can be found in [22]. For an introduction to factor graphs and the SPA, we direct the reader to [11], [23], [24].

II. The Modified NFG $\mathsf{N}_{\mathrm{MB}}(\Theta)$

In this section we introduce the modified NFG $\mathsf{N}_{\mathrm{MB}}(\Theta)$. It is most easily explained by first looking at the case $n = 3$. The following steps are used to obtain $\mathsf{N}_{\mathrm{MB}}(\Theta)$ from $\mathsf{N}_{\mathrm{B}}(\Theta)$ (see also Fig. 1):

1) Fig. 1(a): we define $\mathsf{N}_{\mathrm{B}}(\Theta)$ as in [10]. (See [10, Fig. 1 and Sec. II] for all the details.) Note that the blue function nodes on the left and right encode the entries of Θ.
2) From Fig. 1(a) to Fig. 1(b): we modify the NFG such that green function nodes encode the entries of the matrix Θ. The blue function nodes on the left and the right are now merely suitable indicator functions.
3) From Fig. 1(c to d): for every $i \in [n]$, the function nodes within the ellipse containing $v'_{L,i}$, v'_{ii}, and $v'_{R,i}$ are replaced by a single blue function node w_i in Fig. 1(d). Moreover, for every $1 \leqslant i < j \leqslant n$, the function nodes v'_{ij} and v'_{ji} are merged to the green function node v_{ij} in Fig. 1(d). Finally, pairs of parallel edges are replaced by a single edge and the corresponding variables concatenated.

For general n, the procedure is essentially the same. We start with the NFG $\mathsf{N}_{\mathrm{B}}(\Theta)$ in Fig. 2(left) and obtain the NFG $\mathsf{N}_{\mathrm{MB}}(\Theta)$ in Fig. 2(right). Note that the $\mathsf{N}_{\mathrm{MB}}(\Theta)$ has $\binom{n}{2}$ green function nodes of degree 2 on the left and n blue function nodes of degree n on the right.

Given $\mathsf{N}_{\mathrm{MB}}(\Theta)$, we use the standard approach [11] for formulating the Bethe free energy function F_{MB}. The modified Bethe permanent is then defined to be

$$\mathrm{perm}_{\mathrm{MB}}(\Theta) := \exp\left(- \min_{\underline{b}} F_{\mathrm{MB}}(\underline{b}) \right),$$

where the minimization is over the local marginal polytope associated with $\mathsf{N}_{\mathrm{MB}}(\Theta)$. The quality of the approximation will be measured by

$$\eta_{\mathrm{MB}}(\Theta) := \frac{1}{n} \log \frac{\mathrm{perm}(\Theta)}{\mathrm{perm}_{\mathrm{MB}}(\Theta)}.$$

III. SPA for The Modified NFG $\mathsf{N}_{\mathrm{MB}}(\Theta)$

In this section we first discuss the SPA message update rules for $\mathsf{N}_{\mathrm{MB}}(\Theta)$. Afterwards, we show that a suitable reparameterization leads to noticeable complexity gains.

A. SPA Message Update Rules

Define $\mathcal{W} := \{w_i, i \in [n]\}$ and $\mathcal{V} := \{v_{ij}, 1 \leqslant i < j \leqslant n\}$. For $1 \leqslant i < j \leqslant n$, let ∂w_i denote the set of all neighbors v_{ij} of factor node w_i, i.e., let $\partial w_i := \{v_{ij} : i < j\} \cup \{v_{ji} : j < i\}$, and, likewise, $\partial v_{ij} := \{w_i, w_j\}$. For any neighbouring vertices v, w in the NFG, let $\mu^{(t)}_{v \to w}$ denote the message sent from v to w in iteration t.

The SPA message update equations are as follows: Let $g_{ij}(\underline{a}) := \theta_{ij}^{a_1}\theta_{ji}^{a_2}$. For $\underline{a} \in \{0,1\}^2$, $1 \leqslant i < j \leqslant n$, $w, w' \in \partial v_{ij}$ with $w \neq w'$, and $v := v_{ij}$, we have

$$\mu^{(t)}_{v \to w}(\underline{a}) \propto g_{ij}(\underline{a})\mu^{(t-1)}_{w' \to v}(\underline{a}),$$

where the appropriate proportionality constant ensures that $\sum_{\underline{a}} \mu_{v \to w}(\underline{a}) = 1$. Moreover, for any $i \in [n]$, $w := w_i$, and $v \in \partial w$, we have

$$\mu^{(t)}_{w \to v}(11) \propto \prod_{v' \in \partial w \setminus \{v\}} \mu^{(t)}_{v' \to w}(00),$$

$$\mu^{(t)}_{w \to v}(01) \propto \sum_{v' \in \partial w \setminus \{v\}} \mu^{(t)}_{v' \to w}(10) \prod_{v'' \in \partial w \setminus \{v,v'\}} \mu^{(t)}_{v'' \to w}(00),$$

$$\mu^{(t)}_{w \to v}(10) \propto \sum_{v' \in \partial w \setminus \{v\}} \mu^{(t)}_{v' \to w}(01) \prod_{v'' \in \partial w \setminus \{v,v'\}} \mu^{(t)}_{v'' \to w}(00),$$

$$\begin{aligned}\mu^{(t)}_{w \to v}(00) \propto{}& \theta_{ii} \prod_{v' \in \partial w \setminus \{v\}} \mu^{(t)}_{v \to w}(00) \\ &+ \sum_{v' \in \partial w \setminus \{v\}} \Bigg[\mu^{(t)}_{v' \to w}(11) \prod_{v'' \in \partial w \setminus \{v,v'\}} \mu^{(t)}_{v'' \to w}(00) \\ &+ \mu^{(t)}_{v' \to w}(10) \sum_{v'' \in \partial w \setminus \{v,v'\}} \Bigg(\mu^{(t)}_{v'' \to w}(01) \prod_{v''' \in \partial w \setminus \{v,v',v''\}} \mu^{(t)}_{v''' \to w}(00)\Bigg)\Bigg].\end{aligned}$$

The proportionality constants enforce that $\sum_{\underline{a} \in \{0,1\}^2} \mu_{w \to v}(\underline{a}) = 1$. The messages are updated until convergence, or until some suitable stopping criterion is reached.

B. Improved SPA Message Update Rules

By reparameterizing the messages, we can reduce the number of multiplications required in each iteration, thereby reducing the computational complexity. (This reparameterization is also of relevance for the NFG in [18].) For $\underline{a} \in \{01, 10, 11\}$, and neighbours u, u' in the factor graph, define

$$V^{(t)}_{u \to u'}(\underline{a}) := \mu^{(t)}_{u \to u'}(\underline{a}) / \mu^{(t)}_{u \to u'}(00),$$

The components of the μ-messages sum to 1, hence there is a bijection between the μ messages and the V messages.

In terms of the new messages, we obtain the following message update rules. For every $1 \leqslant i < j \leqslant n$, $\underline{a} \in \{01, 10, 11\}$, $w, w' \in \partial v_{ij}$ and $v := v_{ij}$ we have

$$V^{(t)}_{v \to w}(\underline{a}) := g_{ij}(\underline{a}) V^{(t-1)}_{w' \to v}(\underline{a}).$$

Moreover, for each $i \in [n]$, $w := w_i$, and $v \in \partial w$, we have

$$\begin{aligned}D^{(t)}_{w \to v} :={}& \theta_{ii} + \sum_{v' \in \partial w \setminus \{v\}} V^{(t)}_{v' \to w}(11) \\ &+ \sum_{v', v'' \in \partial w \setminus \{v\}, v' \neq v''} V^{(t)}_{v' \to w}(10) \cdot V^{(t)}_{v'' \to w}(01),\end{aligned}$$

$$V^{(t)}_{w \to v}(11) := \frac{1}{D^{(t)}_{w \to v}},$$

$$V^{(t)}_{w \to v}(10) := \frac{\sum_{v' \in \partial w \setminus \{v\}} V^{(t)}_{v' \to w}(01)}{D^{(t)}_{w \to v}},$$

$$V^{(t)}_{w \to v}(01) := \frac{\sum_{v' \in \partial w \setminus \{v\}} V^{(t)}_{v' \to w}(10)}{D^{(t)}_{w \to v}}.$$

IV. The Bethe permanent of a 3×3 matrix

The $n = 3$ case is rather special as the Bethe free energy function and the SPA algorithm have some nice properties. The NFG is a cycle, illustrated in Fig. 1. In this special case, the SPA message update rules at each node can be represented as linear transformations followed by renormalization. Furthermore, the Bethe entropy function can be written as a sum of conditional entropy functions. We use this to show that the Bethe free energy function is convex.

Viewing the messages $\mu^{(t)}_{v_{ij} \to w_i}, \mu^{(t)}_{w_i \to v_{ij}}$ as column vectors of length 4, we can write $\mu^{(t)}_{w_i \to v_{ij}} \propto A_i \mu^{(t)}_{v_{ij'} \to w_i}$ and $\mu^{(t)}_{v_{ij} \to w_i} \propto A_{ij} \mu^{(t-1)}_{w_j \to v_{ij}}$. Here, for $i = 1, 3$,

$$A_i := \begin{pmatrix} \theta_{ii} & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \; A_2 := \begin{pmatrix} \theta_{22} & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

whereas for $i < j$,

$$A_{ij} := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \theta_{ij} & 0 & 0 \\ 0 & 0 & \theta_{ji} & 0 \\ 0 & 0 & 0 & \theta_{ij}\theta_{ji} \end{pmatrix}.$$

1) Convergence: We can write the recursion

$$\mu^{(t)}_{w_2 \to v_{23}} \propto A_{w_2 \to v_{23}} \mu^{(t-1)}_{w_2 \to v_{23}}$$

where $A_{w_2 \to v_{23}} := A_{23}A_3A_{13}A_1A_{v_{12}}A_2 = (A_2A_{12}A_1A_{13}A_3A_{23})^T$. The message from v_{23} to w_2 can be written as

$$\mu^{(t)}_{v_{23} \to w_2} \propto A_{v_{23} \to w_2} \mu^{(t-1)}_{v_{23} \to w_2} \propto A^T_{w_2 \to v_{23}} \mu^{(t-1)}_{v_{23} \to w_2}$$

Likewise, we can write $\mu^{(t)}_{w_i \to v_{ij}}$ in terms of $\mu^{(t-1)}_{w_i \to v_{ij}}$ as a linear transformation followed by renormalization. Since this factor graph has a single cycle, we can invoke the results of [25], which tell us that the SPA converges and the messages converge to the eigenvector corresponding to the largest eigenvalue of $A_{w_2 \to v_{23}}$, and the partition function is equal to this eigenvalue. We cannot apply directly the Perron-Frobenius theorem to guarantee the existence of a unique largest eigenvalue because the matrices $A_{w_i \to v_{ij}}$ and $A_{v_{ij} \to w_i}$ need not be positive. However, with some minor modifications (permuting rows and columns), we can write $A_{w_i \to v_{ij}}$ as a block diagonal matrix, and try to apply the Perron-Frobenius theorem to each block. Yet, the existence of a unique largest eigenvalue for the matrix is not guaranteed. If the largest eigenvalue has multiplicity greater than one, then the SPA might oscillate. However, these scenarios can typically be circumvented by damping the messages.[2] The SPA is guaranteed to converge if the initial messages have nonzero projection onto the largest eigenvector. If we choose the initial vector randomly, then it will have a nonzero component in the direction of the eigenvector for the largest eigenvalue with high probability. We can prove the following.

Proposition 1. *If the messages are initialized at random according to any distribution with full support over the probability simplex, then the SPA converges with probability* 1*, and* $\mathrm{perm}_{\mathrm{MB}}(\Theta)$ *is equal to the largest eigenvalue of* $A_{w_2 \to v_{23}}$.

Proof. See [22]. □

2) Convexity: The Bethe entropy function can be written as a sum of conditional entropy functions, which in turn is a concave function of the joint distribution. This lets us conclude that the Bethe free energy function is convex.

Proposition 2. *The Bethe entropy function is a concave function of the beliefs. Hence, the Bethe free energy function is convex.*

Proof. See [22]. □

3) Correctness: We can prove the following proposition by explicitly computing an expression for the largest eigenvalue and comparing it with the permanent. Moreover, the bounds for η_{MB} are tight.

Proposition 3. *For every* 3×3 *nonnegative matrix* Θ *with a nonzero permanent,* $0 \leqslant \eta_{\mathrm{MB}}(\Theta) \leqslant \frac{1}{3}$.

Proof. See [22]. □

V. The General Case

A. Convexity of the Bethe free energy function

We now proceed to study the case when $n > 3$. In this scenario, the Bethe entropy function is not a concave function of the beliefs. In fact, we show in [22] that there exist vertices $\underline{b}_v$ in the belief polytope $\mathcal{B}$, direction $\underline{\xi}$ and a small $\epsilon > 0$ such that $H_{\mathrm{B}}(\underline{b}_v + t\underline{\xi})$ is a convex function of t for $t \in [0, \epsilon]$. The proof follows by computing expressions for the second directional derivative, and showing the existence of $\underline{\xi}$ for which this is equal to $+\infty$.

Proposition 4. *For* $n > 3$*,* F_{MB} *is not convex.*

Proof. See [22]. □

Although there are directions along which F_{MB} is concave, simulations lead us to conjecture that for every point in $\mathcal{B}$, there are always directions along which F_{MB} is convex, and that F_{MB} does not have non-global local minima.

[2]This is generally done by updating the messages using $\mu^{(t)} \leftarrow \alpha\mu^{(t)} + (1-\alpha)\mu^{(t-1)}$ at the end of each iteration, where α is the damping constant and is chosen heuristically.

B. Vertices of the local marginal polytope

The structure of the local marginal polytope can give us insight into the behavior of the SPA and the Bethe free energy function. Let us call a point $\underline{b}$ in $\mathcal{B}$ integral if all the entries of $\underline{b}$ are either 0 or 1. Likewise, $\underline{b} \in \mathcal{B}$ is called fractional if there is some entry of $\underline{b}$ which is strictly between 0 and 1. Clearly, every integral point in $\mathcal{B}$ is a vertex. In fact, it is also easy to see that every integral point corresponds to a permutation. We initially suspected that $\mathcal{B}$ contains no fractional vertices, but a numerical search showed that this is not true. See [22] for more details.

C. An exact expression for the approximation ratio

The expression in the following lemma is of interest because of the permanent expression in the numerator of the fraction.

Lemma 1. *For any stationary point* $\underline{b}$ *of the Bethe free energy function, we have*

$$\eta_{\mathrm{MB}}(\Theta) = \frac{1}{n} \log \left(\frac{\sum_{\sigma \in S_n} \prod_i \varphi_i(\sigma)}{\prod_i \prod_{j \in [n] \setminus \{i\}} \sqrt{b_{00}^{(ij)}}} \right),$$

where

$$\varphi_i(\sigma) = \begin{cases} b_{ii}^{(i)}, & \text{if } \sigma(i) = i \\ \sqrt{b_{00}^{(ij)} b_{11}^{(ij)}}, & (\sigma(i), \sigma(j)) = (j, i), \text{ for } j \neq i \\ b_{lm}^{(i)} \sqrt{\frac{b_{00}^{(il)} b_{00}^{(im)}}{b_{10}^{(il)} b_{01}^{(im)}}}, & (\sigma(i), \sigma(m)) = (l, i), \\ & \text{for distinct } l, m, i. \end{cases}$$

Proof. See [22]. □

VI. Simulation Results

In our simulations, we first run a maximum-weight matching algorithm on a randomly generated matrix Θ and then permute the columns to give a Θ^{pre} (superscript indicates preprocessed) for which $\prod_i \theta_{i\sigma(i)}^{\mathrm{pre}}$ is maximized by the identity permutation σ. Fig. 3 illustrates that this preprocessing greatly improves the performance of the algorithm.

The main results are tabulated in Tables I and II. We observe that the proposed algorithm performs particularly well when Θ is close to being block diagonal. In all our simulations, we observed that the proposed algorithm converges (after introducing suitable damping). We observed that for even moderately large n, the value η_{MB} is much smaller than $1/2$. We ran a numerical search for the worst matrix that maximizes η_{MB}, but only obtained values very close to those in Table I. We also performed the simulations with initial messages corresponding to integral vertices of the local marginal polytope, and observed that the SPA converges to an internal point in the polytope even if convergence is slowed by damping. We saw in Sec. V-A that for certain integral vertices, there exist directions along which the Bethe entropy function has infinite slope. This suggests that even though the Bethe free energy function has "steep hills" near certain vertices, there are also "valleys" along which the SPA can escape these hills to move towards the interior of the polytope. We conjecture that for every point in the local marginal polytope, there exist directions along which the Bethe free energy function is convex, and that F_{MB} can have multiple local maxima but has no non-global local minima.

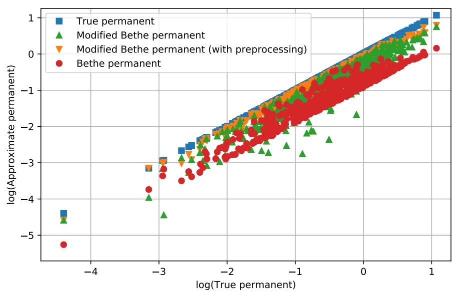

Fig. 3: Simulation results for $n = 3$. Entries of the random matrix were generated uniformly at random from $[0, 1]$. Here, preprocessing refers to the permutation of the rows/columns using the max-weight matching algorithm. The preprocessing significantly improves the performance of our algorithm.

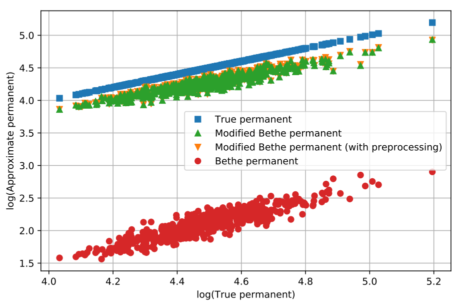

Fig. 4: Simulation results for $n = 12$. The random matrix Θ is obtained by adding random noise to the tensor product of $1_{2\times 2}$ and I_6. Here, preprocessing refers to the permutation of the rows/columns using the maximum-weight matching algorithm.

TABLE I: Simulation results for 12×12 block-diagonal matrices. Matrices were generated by adding random i.i.d. noise to the tensor product $1_{k\times k} \otimes I_{12/k}$. For each k, results were obtained by averaging over 1000 matrices.

k	$\mathbb{E}\eta_{\mathrm{MB}}(\Theta)$	$\mathbb{E}\eta_{\mathrm{B}}(\Theta)$	$\max \eta_{\mathrm{MB}}(\Theta)$	$\max \eta_{\mathrm{B}}(\Theta)$
2	0.003	0.42	0.004	0.43
3	0.16	0.38	0.17	0.39
4	0.28	0.34	0.28	0.35
6	0.27	0.28	0.28	0.28
12	0.20	0.20	0.20	0.20

TABLE II: Simulation results for diagonally dominant matrices obtained by adding a diagonal matrix to the adjacency matrix of a random Erdős-Rényi graph. For each k, results were obtained by averaging over 1000 matrices.

k	$\mathbb{E}\eta_{\mathrm{MB}}(\Theta)$	$\mathbb{E}\eta_{\mathrm{B}}(\Theta)$	$\max \eta_{\mathrm{MB}}(\Theta)$	$\max \eta_{\mathrm{B}}(\Theta)$
12	0.01	0.35	0.08	0.10
13	0.02	0.04	0.08	0.10
15	0.03	0.06	0.08	0.10

References

[1] M. Chertkov, L. Kroc, and M. Vergassola, "Belief propagation and beyond for particle tracking," *arXiv:0806.1199*, 2008.
[2] P. O. Vontobel, "The Bethe approximation of the pattern maximum likelihood distribution," in *2012 IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, 2012.
[3] H. J. Ryser, *Combinatorial Mathematics*. Math. Assoc. America; dist. Wiley [New York], 1963, no. 14.
[4] L. G. Valiant, "The complexity of computing the permanent," *Theoretical Comp. Sc.*, vol. 8, no. 2, pp. 189–201, 1979.
[5] N. Karmarkar, R. Karp, R. Lipton, L. Lovász, and M. Luby, "A Monte-Carlo algorithm for estimating the permanent," *SIAM J. Computing*, vol. 22, no. 2, pp. 284–293, 1993.
[6] A. Barvinok, "Polynomial time algorithms to approximate permanents and mixed discriminants within a simply exponential factor," *Random Structures & Algorithms*, vol. 14, no. 1, pp. 29–61, 1999.
[7] N. Linial, A. Samorodnitsky, and A. Wigderson, "A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents," *Combinatorica*, vol. 20, no. 4, pp. 545–568, 2000.
[8] M. Jerrum, A. Sinclair, and E. Vigoda, "A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries," *J. ACM*, vol. 51, no. 4, pp. 671–697, 2004.
[9] B. Huang and T. Jebara, "Approximating the permanent with belief propagation," *arXiv:0908.1769*, 2009.
[10] P. O. Vontobel, "The Bethe permanent of a nonnegative matrix," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1866–1901, Mar. 2013.
[11] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2288–2312, Jul. 2005.
[12] L. Gurvits, "Unleashing the power of Schrijver's permanental inequality with the help of the Bethe approximation," *arXiv preprint arXiv:1106.2844*, 2011.
[13] D. Straszak and N. K. Vishnoi, "Belief propagation, Bethe approximation and polynomials," in *Proc. Allerton Conf. Comm., Cont., and Computing*, Monticello, IL, USA, Oct. 2017, pp. 666–671.
[14] L. Gurvits and A. Samorodnitsky, "Bounds on the permanent and some applications," in *Proc. Ann. Symp. Found. Comp. Science*, PA, USA, Oct. 2014, pp. 90–99.
[15] N. Anari and A. Rezaei, "A tight analysis of bethe approximation for permanent," in *Proc. Ann. Symp. Found. Comp. Science*, Baltimore, MD, USA, Nov. 2019, pp. 1434–1445.
[16] M. Bayati, D. Shah, and M. Sharma, "Max-product for maximum weight matching: Convergence, correctness, and LP duality," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, p. 12411251, Mar. 2008.
[17] S. Sanghavi, D. Malioutov, and A. Willsky, "Belief propagation and LP relaxation for weighted matching in general graphs," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2203–2212, Apr. 2011.
[18] M. X. Cao and P. O. Vontobel, "Double-edge factor graphs: definition, properties, and examples," in *Proc. Inf. Theory Workshop*, Kaohsiung, Taiwan, Nov. 2017, pp. 136–140.
[19] H.-A. Loeliger and P. O. Vontobel, "Factor graphs for quantum probabilities," *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 5642–5665, Sep. 2017.
[20] ——, "Quantum measurement as marginalization and nested quantum systems," *IEEE Trans. Inf. Theory, to appear*, 2020.
[21] M. X. Cao and P. O. Vontobel, "Bounding and estimating the classical information rate of quantum channels with memory," *IEEE Trans. Inf. Theory, to appear*, 2020.
[22] S. Vatedka and P. O. Vontobel, "Modified Bethe permanent of a nonnegative matrix," *preprint*, 2020. [Online]. Available: www.iith.ac.in/ shashankvatedka/html/publications.html
[23] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Sig. Proc. Magazine*, vol. 21, no. 1, pp. 28–41, Jan. 2004.
[24] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
[25] Y. Weiss, "Correctness of local probability propagation in graphical models with loops," *Neural Computation*, vol. 12, no. 1, pp. 1–41, 2000.