Proc. 48th Allerton Conf. on Communications, Control, and Computing,
Allerton House, Monticello, IL, USA, Sep. 29 - Oct. 1, 2010.

The Bethe Permanent of a Non-Negative Matrix*

Pascal O. Vontobel

***Abstract*— It has recently been observed that the permanent of a non-negative matrix, i.e., of a matrix containing only non-negative real entries, can very well be approximated by solving a certain Bethe free energy minimization problem with the help of the sum-product algorithm. We call the resulting approximation of the permanent the Bethe permanent.**

In this paper we give reasons why this approach to approximating the permanent works well. Namely, we show that the Bethe free energy is a convex function and that the sum-product algorithm finds its minimum efficiently. We also show that the permanent is lower bounded by the Bethe permanent, and we list some empirical evidence that the permanent is upper bounded by some constant (that modestly grows with the matrix size) times the Bethe permanent. Part of these results are obtained by a combinatorial characterization of the Bethe permanent in terms of permanents of so-called lifted versions of the matrix under consideration.

We conclude the paper with some conjectures about permanent-based pseudo-codewords and permanent-based kernels, and we comment on possibilities to modify the Bethe permanent so that it approximates the permanent even better.

I. Introduction

Central to the topic of this paper is the definition of the permanent of a square matrix (see, e.g., [1]).

Definition 1 *Let $\boldsymbol{\theta} = (\theta_{i,j})_{i,j}$ be a real matrix of size $n \times n$. The permanent of $\boldsymbol{\theta}$ is defined to be the scalar*

$$\operatorname{perm}(\boldsymbol{\theta}) = \sum_{\sigma} \prod_{i \in [n]} \theta_{i,\sigma(i)}, \qquad (1)$$

where the summation is over all $n!$ permutations of the set $[n] \triangleq \{1, 2, \ldots, n\}$. □

Contrast this definition with the definition of the *determinant* of $\boldsymbol{\theta}$, i.e.,

$$\det(\boldsymbol{\theta}) = \sum_{\sigma} \operatorname{sgn}(\sigma) \prod_{i \in [n]} \theta_{i,\sigma(i)},$$

where $\operatorname{sgn}(\sigma)$ equals $+1$ if σ is an even permutation and equals -1 if σ is an odd permutation.

A. Complexity of Computing the Permanent

Because the definition of the permanent looks simpler than the definition of the determinant, it is tempting to conclude that the permanent can be computed at least as efficiently as the determinant. However, this does not seem to be the case. Namely, whereas the arithmetic complexity (number of real additions and multiplications) needed to compute the determinant is in $O(n^3)$, Ryser's algorithm (one of the best known algorithms for computing the permanent) requires $\Theta(n \cdot 2^n)$ arithmetic operations [2]. This clearly improves upon the brute-force complexity $O(n \cdot n!) = O\big(n^{3/2} \cdot (n/e)^n\big)$ for computing the permanent, but is still exponential in the matrix size.

In terms of complexity classes, the computation of the permanent is in the complexity class #P ("sharp P" or "number P") [3], where #P is the set of the counting problems associated with the decision problems in the set NP. Note that even the computation of the permanent of zero-one matrices is #P-complete. Therefore, the above-mentioned complexity numbers for the computation of the permanent are not surprising.

B. Approximations to the Permanent

Given the difficulty of computing the permanent exactly, and given the fact that in many applications it is good enough to compute an *approximation* to the permanent, this paper focuses on efficient methods to approximate the permanent. This relaxation in requirements, from exact to approximate evaluation of the permanent, allows one to devise algorithms that potentially have much lower complexity.

Moreover, we will consider only the case where the matrix $\boldsymbol{\theta}$ in (1) is non-negative, i.e., where all entries of $\boldsymbol{\theta}$ are non-negative. It is to be expected that approximating the permanent is simpler in this case because with this restriction the sum in (1) contains only non-negative terms, i.e., the terms in this sum "interfere constructively." This is in contrast to the general case where the sum in (1) contains positive and negative terms, i.e., the terms in this sum "interfere constructively and destructively." Despite this restriction to non-negative matrices, many interesting counting problems can be captured by this setup.

Earlier work on approximating the permanent of a non-negative matrix includes Markov-chain-Monte-Carlo-based methods by Broder (see [4]), fully polynomial-time randomized approximation schemes (FPRAS) [5], and Bethe-approximation-based / sum-product-algorithm-based methods [6], [7].

The study in this paper was very much motivated by this last set of papers on graphical-model-based methods, in particular by the fact these methods yield algorithms that are *very efficient* and by the fact that the obtained permanent estimates have an *accuracy that is good enough for many purposes*.

The idea behind this graphical-model-based approach is to formulate a factor graph whose partition function equals the

P. O. Vontobel is with Hewlett–Packard Laboratories, 1501 Page Mill Road, Palo Alto, CA 94304, USA (email: pascal.vontobel@ieee.org).

*Due to space limitations, some of the text and the proofs have been omitted. They will appear as part of a longer paper on the topic of this paper.

Proc. 48th Allerton Conf. on Communications, Control, and Computing,
Allerton House, Monticello, IL, USA, Sep. 29 - Oct. 1, 2010.

permanent that we are looking for. Consequently, the negative logarithm of this partition function can be written as the minimum of the so-called Gibbs free energy function that is associated with this factor graph. Although being an elegant reformulation of the permanent computation problems, this does not yet yield any computational savings. Nevertheless, it suggests to look for a function that is tractable and whose minimum is close to the minimum of the Gibbs free energy. One such function is the so-called Bethe free energy function [8]. The Bethe free energy function is interesting because a theorem by Yedidia, Freeman, and Weiss [8] says that fixed points of the sum-product algorithm correspond to stationary points of the Bethe free energy.

In general, this approach of replacing the Gibbs free energy function by the Bethe free energy function comes with very few guarantees, though.

- The Bethe free energy function might have multiple local minima.
- It is unclear how close the (global) minimum of the Bethe free energy is to the minimum of the Gibbs free energy.
- It is unclear if the sum-product algorithm converges (even to a local minimum of the Bethe free energy).

Luckily, in the case of the permanent approximation problem, one can formulate a factor graph where the Bethe free energy function is very well behaved. In particular, in this paper we show that a factor graph with the following properties can be formulated.

- The Bethe free energy function is *convex* and therefore has no local minima.
- The minimum of the Bethe free energy is *quite close* to the minimum of the Gibbs free energy. (In particular, the permanent is lower bounded by the Bethe permanent. Moreover, empirical evidence suggests that the permanent is upper bounded by some constant (that modestly grows with the matrix size) times the Bethe entropy.)
- The sum-product algorithm *converges* to the minimum of the Bethe free energy. (In fact, the error w.r.t. minimum of the Bethe free energy decays exponentially fast, with an exponent depending on the matrix $\boldsymbol{\theta}$.)

Besides leaving some questions open with respect to the Bethe free energy function (see for example the above-mentioned conjecture), these results by-and-large validate the empirical success, as observed by Chertkov *et al.* [6] and by Huang and Jebara [7], of approximating the permanent by graphical-model-based methods.[1]

As we will see, the factor graph that we use (cf. Figure 1) is not sparse and has many short cycles, in particular many four-cycles. These facts might suggest that the application of the sum-product algorithm to this factor graph is rather problematic. However, as we have discussed above, and will show in this paper, luckily this is not the case.

[1] Note that the convexity of a function in general, and of the Bethe free energy in particular, depends on its parametrization. That might explain the non-convexity observations in [7, Section 3.3] w.r.t. the Bethe free energy.

C. Related Work

Let us briefly mention some related work. Namely, some aspects of the Bethe free energy function were analyzed by Watanabe and Chertkov in [9] (in particular, they applied the loop calculus technique by Chernyak and Chertkov) and by Chertkov et al. in [10].

Because computing the permanent is related to counting perfect matchings, the paper by Bayati and Nair [11] (of which we became aware of while writing up the present paper) on counting matchings in graphs with the help of the sum-product algorithm is very relevant. Note that their setup is such that the perfect matching case can be seen as a limiting case of the matching setup. However, for the perfect matching case (a case for which the authors of [11] make no claims) the convergence proof of the sum-product algorithm in [11] is incomplete.

D. Overview of Paper

This paper is structured as follows. We conclude this introductory section with a discussion of some of the notation that is used. In Section II we then introduce the main normal factor graph for this paper, in Section III we formally define the Bethe permanent, in Section IV we discuss properties of the Bethe entropy function and the Bethe free energy function, in Section V we analyze the sum-product algorithm, in Section VI we give a “combinatorial characterization” of the Bethe permanent, in Section VII we bound the permanent in terms of the Bethe permanent, in Section VIII we list some thoughts on using the concept of “fractional Bethe entropy,” and in Section IX we list some conjectures.

E. Basic Notations and Definitions

This subsection discusses the most important notations that will be used in this paper. More notational definitions will be given in later sections.

We let $\mathbb{R}$ be the field of real numbers. Scalars are denoted by non-boldface characters, whereas vectors and matrices by boldface characters. For any positive integer L, the matrix $\mathbf{1}_{L\times L}$ is the all-one matrix of size $L \times L$.

We use calligraphic letters for sets, and the size of a set $\mathcal{S}$ is denoted by $|\mathcal{S}|$. The convex hull [12] of some subset $\mathcal{S}$ of some multi-dimensional real space is denoted by $\mathrm{conv}(\mathcal{S})$. For any positive integer L we define $[L] \triangleq \{1, \ldots, L\}$. For any positive integer L, we define $\mathcal{P}_{L\times L}$ to be the set of all $L \times L$ permutation matrices, i.e.,

$$\mathcal{P}_{L\times L} \triangleq \left\{ \boldsymbol{T} \;\middle|\; \begin{array}{l} \boldsymbol{T} \text{ is a matrix of size } L \times L \\ \boldsymbol{T} \text{ contains exactly one } 1 \text{ per row} \\ \boldsymbol{T} \text{ contains exactly one } 1 \text{ per column} \\ \boldsymbol{T} \text{ contains } 0\text{s otherwise} \end{array} \right\},$$

Clearly, there is a bijection between $\mathcal{P}_{L\times L}$ and the set of all permutations of $[L]$. Moreover, for a finite set $\mathcal{S}$, we define $\Pi_{\mathcal{S}}$ to be the set of probability mass functions over $\mathcal{S}$, i.e.,

$$\Pi_{\mathcal{S}} \triangleq \left\{ \boldsymbol{p} = (p_s)_{s\in\mathcal{S}} \;\middle|\; p_s \geqslant 0 \text{ for all } s \in \mathcal{S},\ \sum_{s\in\mathcal{S}} p_s = 1 \right\}.$$

Proc. 48th Allerton Conf. on Communications, Control, and Computing,
Allerton House, Monticello, IL, USA, Sep. 29 - Oct. 1, 2010.

Finally, for any positive integer L, we let $\Delta_{L\times L}$ be the set of doubly stochastic matrices of size $L \times L$, i.e.,

$$\Delta_{L\times L} \triangleq \left\{ \boldsymbol{\gamma} = \left(\gamma_{i,j}\right) \,\middle|\, \begin{array}{l} \gamma_{i,j} \geqslant 0 \text{ for all } (i,j) \in [L]\times[L] \\ \sum_{j\in[L]} \gamma_{i,j} = 1 \text{ for all } i \in [L] \\ \sum_{i\in[L]} \gamma_{i,j} = 1 \text{ for all } j \in [L] \end{array} \right\}.$$

Clearly, $\Delta_{L\times L}$ is a convex set, and every permutation matrix of size $L \times L$ is a doubly stochastic matrix of size $L \times L$. Most interestingly, every doubly stochastic matrix of size $L\times L$ can be written as a convex combination of permutation matrices of size $L \times L$; this observation is a consequence of the important Birkhoff–von Neumann Theorem.

Theorem 2 (Birkhoff–von Neumann Theorem) *For any positive integer L, the set of doubly stochastic matrices of size $L \times L$ is a polytope whose vertex set equals the set of permutation matrices of size $L \times L$, i.e.,*

$$\text{vertex-set}(\Delta_{L\times L}) = \mathcal{P}_{L\times L}.$$

As a consequence, the set of doubly stochastic matrices of size $L \times L$ is the convex hull of the set of all permutation matrices of size $L \times L$, i.e.,

$$\Delta_{L\times L} = \text{conv}(\mathcal{P}_{L\times L}).$$

Proof: See, for example, [13, Section 8.7]. □

II. Normal Factor Graph Representation

Factor graphs are a convenient way to represent multivariate functions [14]. In this paper we use a variant called “normal factor graphs” [15] (also called “Forney-style factor graphs” [16]), where variables are associated with edges.

As already mentioned in the introduction, the main idea behind the graphical-model-based approach to estimating the permanent is to formulate a normal factor graph such that the partition function equals the permanent. There are of course different ways to do this and typically different formulations will yield different results when estimating the permanent with sub-optimal algorithms like the sum-product algorithm. It is well known that when the normal factor graph has no cycles, then we get the permanent exactly, however, for the given problem any normal factor graph *without* cycles yields highly inefficient sum-product update rules (for reasonably large n), and so we will focus on normal factor graphs *with* cycles. The normal factor graph that is introduced in the following definition and that is based on a complete bipartite graph with two times n vertices is a rather natural candidate, and as we will see, has very interesting and useful properties.

Throughout the paper, if not mentioned otherwise, $\boldsymbol{\theta}$ will represent a non-negative matrix of size $n \times n$, where n is some fixed positive integer.

Definition 3 *We define the normal factor graph $\mathsf{N}(\boldsymbol{\theta}) \triangleq \mathsf{N}(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ as follows (see also Figure 1).*

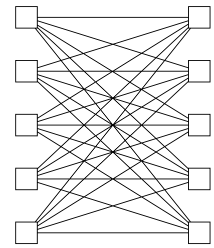

Fig. 1. The normal factor graph $\mathsf{N}(\boldsymbol{\theta})$ which is based on a complete bipartite graph with two times n vertices. The left function nodes represent the functions $\{g_i\}_{i\in\mathcal{I}}$, the right function nodes represent the functions $\{g_j\}_{j\in\mathcal{J}}$, and with edge $e = (i,j)$ we associate the variable label $A_e = A_{i,j}$. (See Definition 3 for more details.)

- *The set of vertices is $\mathcal{F} \triangleq \mathcal{I}\dot{\cup}\mathcal{J}$ where $\mathcal{I} \triangleq [n]$ will be called the set of left vertices and $\mathcal{J} \triangleq [n]$ will be called the set of right vertices.*[2]
- *The set of edges is $\mathcal{E} \triangleq \mathcal{I}\times\mathcal{J} = \big\{(i,j) \,\big|\, i \in \mathcal{I}, j \in \mathcal{J}\big\}$. Moreover, all edges are full edges, i.e., $\mathcal{E}_{\text{full}} = \mathcal{E}$ and $\mathcal{E}_{\text{half}} = \{\,\}$.*
- *For every $e = (i,j) \in \mathcal{E}$ we define $\mathcal{A}_e \triangleq \mathcal{A}_{i,j} \triangleq \{0,1\}$.*
- *For every $i \in \mathcal{I}$ we define the local functions*[3]

$$g_i : \prod_j \mathcal{A}_{i,j} \to \mathbb{R}, \quad \boldsymbol{a}_i \mapsto \begin{cases} \sqrt{\theta_{i,j}} & \text{(if } \boldsymbol{a}_i = \boldsymbol{e}_j\text{)} \\ 0 & \text{(otherwise)} \end{cases}$$

Similarly, for every $j \in \mathcal{J}$ we define the local functions

$$g_j : \prod_i \mathcal{A}_{i,j} \to \mathbb{R}, \quad \boldsymbol{a}_j \mapsto \begin{cases} \sqrt{\theta_{i,j}} & \text{(if } \boldsymbol{a}_j = \boldsymbol{e}_i\text{)} \\ 0 & \text{(otherwise)} \end{cases}$$

- *For every $i \in \mathcal{I}$ we define the function node alphabet $\mathcal{A}_i$ to be the set*

$$\mathcal{A}_i \triangleq \left\{ \boldsymbol{a}_i \in \prod_j \mathcal{A}_{i,j} \,\middle|\, g_i(\boldsymbol{a}_i) \neq 0 \right\} = \{\boldsymbol{e}_j \mid j \in \mathcal{J}\}\,.$$

Similarly, for every $j \in \mathcal{J}$ we define the function node alphabet $\mathcal{A}_j$ to be the set

$$\mathcal{A}_j \triangleq \left\{ \boldsymbol{a}_j \in \prod_i \mathcal{A}_{i,j} \,\middle|\, g_j(\boldsymbol{a}_j) \neq 0 \right\} = \{\boldsymbol{e}_i \mid i \in \mathcal{I}\}\,.$$

- *With a slight abuse of notation, the letter $\mathcal{A}$ will also denote the set $\mathcal{A} \triangleq \prod_e \mathcal{A}_e = \prod_{i,j} \mathcal{A}_{i,j}$. With this, a vector*

$$\boldsymbol{a} \;\triangleq\; (a_e)_{e\in\mathcal{E}} = (a_{i,j})_{(i,j)\in\mathcal{I}\times\mathcal{J}} \;\in\; \mathcal{A}$$

will be called a configuration of the normal factor graph. For a given vector $\boldsymbol{a}$, we also define the sub-vectors

$$\boldsymbol{a}_i \triangleq \big\{(a_{i,j})_{j\in\mathcal{J}}\big\} \quad \text{and} \quad \boldsymbol{a}_j \triangleq \big\{(a_{i,j})_{i\in\mathcal{I}}\big\}.$$

[2] Here, $\mathcal{F} \triangleq \mathcal{I}\dot{\cup}\mathcal{J}$ stands for the more cumbersome $\mathcal{F} \triangleq \big(\{\text{left}\}\times\mathcal{I}\big) \cup \big(\{\text{right}\}\times\mathcal{J}\big)$. In the following, i (and variations thereof) will refer to a left vertex and j (and variations thereof) will refer to a right vertex. In that spirit, variables like η_i and η_j are different variables, also if $i = j$.

[3] Here and in the following, $\boldsymbol{e}_j$, $j \in \mathcal{J}$, stands for the length-n vector where all entries are zero except for the j-th entry that equals 1. The vector $\boldsymbol{e}_i$, $i \in \mathcal{I}$, is defined similarly.

Proc. 48th Allerton Conf. on Communications, Control, and Computing,
Allerton House, Monticello, IL, USA, Sep. 29 - Oct. 1, 2010.

When convenient, the vector $\boldsymbol{a}$ will be considered to be an $n \times n$ matrix. Then $\boldsymbol{a}_i$ corresponds to the i-th row of $\boldsymbol{a}$, and $\boldsymbol{a}_j$ corresponds to the j-th column of $\boldsymbol{a}$.

- *The global function g is defined to be*

$$g: \mathcal{A} \to \mathbb{R}, \quad \boldsymbol{a} \mapsto \left(\prod_i g_i(\boldsymbol{a}_i) \right) \cdot \left(\prod_j g_j(\boldsymbol{a}_j) \right).$$

- *A configuration $\boldsymbol{c}$ with $g(\boldsymbol{c}) \neq 0$ will be called a valid configuration. Clearly, the set of valid configurations $\mathcal{C}_{\mathrm{edge}}$ is characterized as follows*

$$\mathcal{C}_{\mathrm{edge}} \triangleq \left\{ (c_{i,j})_{i,j \in \mathcal{I} \times \mathcal{J}} \;\middle|\; \begin{array}{c} c_{i,j} \in \mathcal{A}_{i,j},\, (i,j) \in \mathcal{I} \times \mathcal{J} \\ \boldsymbol{c}_i \in \mathcal{A}_i,\, i \in \mathcal{I} \\ \boldsymbol{c}_j \in \mathcal{A}_j,\, j \in \mathcal{J} \end{array} \right\}.$$

Considering $\boldsymbol{c} \in \mathcal{C}_{\mathrm{edge}}$ as an $n \times n$ matrix, it can easily be verified that $\boldsymbol{c}$ is a permutation matrix, i.e., $\boldsymbol{c} \in \mathcal{P}_{n \times n}$.

In the following, we denote by F_{G}, U_{G}, H_{G}, and Z_{G} the Gibbs free energy function, the Gibbs average energy function, the Gibbs entropy function, and the Gibbs partition function, respectively, associated with a normal factor graph. (In the following, when confusion can arise what normal factor graph a certain Gibbs partition function is referring to, we will use $Z_{\mathrm{G}}(\mathsf{N})$, etc., instead of Z_{G}.)

Lemma 4 *Consider the normal factor graph $\mathsf{N}(\boldsymbol{\theta})$. The permanent of $\boldsymbol{\theta}$ can be expressed in terms of the Gibbs partition function or in terms of the minimum of the Gibbs free energy function. Namely,*

$$\mathrm{perm}(\boldsymbol{\theta}) = Z_{\mathrm{G}} = \exp\left(-\min_{\boldsymbol{p}} F_{\mathrm{G}}(\boldsymbol{p}) \right),$$

where the minimization is over $\boldsymbol{p} \in \Pi_{\mathcal{A}}$.

Proof: Omitted. □

Although the reformulation of the permanent in the above lemma is elegant, from a computational perspective it does not buy us much. However, it suggest to look for an optimization problem that can easily be minimized and whose minimal value is related to the desired quantity. This is the approach that is taken in the next section.

III. The Bethe Permanent

For a motivation of the Bethe free energy we refer to Section I-B.

Definition 5 *Consider the normal factor graph $\mathsf{N}(\boldsymbol{\theta})$. We let*

$$\boldsymbol{\beta} \triangleq \big((\boldsymbol{\beta}_i)_{i \in \mathcal{I}}, (\boldsymbol{\beta}_j)_{j \in \mathcal{J}}, (\boldsymbol{\beta}_e)_{e \in \mathcal{E}} \big)$$

be a collection of vectors based on the real vectors

$$\boldsymbol{\beta}_i \triangleq (\beta_{i,\boldsymbol{a}_i})_{\boldsymbol{a}_i \in \mathcal{A}_i},$$
$$\boldsymbol{\beta}_j \triangleq (\beta_{j,\boldsymbol{a}_j})_{\boldsymbol{a}_j \in \mathcal{A}_j},$$
$$\boldsymbol{\beta}_e \triangleq (\beta_{e,a_e})_{a_e \in \mathcal{A}_e}.$$

Moreover, we define the sets

$$\mathcal{B}_i \triangleq \Pi_{\mathcal{A}_i}, \quad i \in \mathcal{I},$$
$$\mathcal{B}_j \triangleq \Pi_{\mathcal{A}_j} \quad j \in \mathcal{J},$$
$$\mathcal{B}_e \triangleq \Pi_{\mathcal{A}_e} \quad e \in \mathcal{E}.$$

and call $\mathcal{B}_i$, $\mathcal{B}_j$, and $\mathcal{B}_e$, the ith local marginal polytope, the jth local marginal polytope, the eth local marginal polytope, respectively. (Sometimes $\mathcal{B}_i$ is also called the ith belief polytope, etc.)

With this, the local marginal polytope (or belief polytope) $\mathcal{B}$ is defined to be the set

$$\mathcal{B} = \left\{ \boldsymbol{\beta} \;\middle|\; \begin{array}{c} \boldsymbol{\beta}_i \in \mathcal{B}_i \text{ for all } i \in \mathcal{I} \\ \boldsymbol{\beta}_j \in \mathcal{B}_j \text{ for all } j \in \mathcal{J} \\ \boldsymbol{\beta}_e \in \mathcal{B}_e \text{ for all } e \in \mathcal{E} \\ \sum_{\boldsymbol{a}'_i \in \mathcal{A}_i :\, a'_{i,e} = a_e} \beta_{i,\boldsymbol{a}'_i} = \beta_{e,a_e} \\ \text{for all } e = (i,j) \in \mathcal{E},\, a_e \in \mathcal{A}_e \\ \sum_{\boldsymbol{a}'_j \in \mathcal{A}_j :\, a'_{j,e} = a_e} \beta_{j,\boldsymbol{a}'_j} = \beta_{e,a_e} \\ \text{for all } e = (i,j) \in \mathcal{E},\, a_e \in \mathcal{A}_e \end{array} \right\},$$

where $\boldsymbol{\beta} \in \mathcal{B}$ is called a pseudo-marginal. (The two constraints that were listed last in the definition of $\mathcal{B}$ will be called "edge consistency constraints.") □

With this, we define the Bethe free energy function F_{B}, the Bethe average energy function U_{B}, and the Bethe entropy function H_{B} associated with a (normal) factor graph as in [8].

Definition 6 *The Bethe partition function associated with some normal factor graph $\mathsf{N}(\mathcal{F}, \mathcal{E}, \mathcal{A}, \mathcal{G})$ is defined to be*

$$Z_{\mathrm{B}} \triangleq \exp\left(-\min_{\boldsymbol{\beta} \in \mathcal{B}} F_{\mathrm{B}}(\boldsymbol{\beta}) \right).$$

In the following, when confusion can arise what normal factor graph a certain Bethe partition function is referring to, we will use $Z_{\mathrm{B}}(\mathsf{N})$, etc., instead of Z_{B}.

The next definition is the main definition of this paper and was motivated by the work of Chertkov et al. [6] and by the work of Huang and Jebara [7].

Definition 7 *Consider the normal factor graph $\mathsf{N}(\boldsymbol{\theta})$. The Bethe permanent of $\boldsymbol{\theta}$, which will be denoted by $\mathrm{perm}_{\mathrm{B}}(\boldsymbol{\theta})$, is defined to be*

$$\mathrm{perm}_{\mathrm{B}}(\boldsymbol{\theta}) \triangleq Z_{\mathrm{B}},$$

where Z_{B} is specified in Definition 6.

IV. Properties of the Bethe Entropy Function and the Bethe Free Energy Function

The Bethe entropy function is in general not a concave function and therefore the Bethe free energy function is in general not a convex function. (Note that the Bethe average energy function is linear in the arguments.) However, here we show that the Bethe entropy function, when suitably

Proc. 48th Allerton Conf. on Communications, Control, and Computing,
Allerton House, Monticello, IL, USA, Sep. 29 - Oct. 1, 2010.

parametrized, is a concave function. Consequently the Bethe free energy, when suitably parametrized, is a convex function. As a consequence, the Bethe free energy does *not* have local minima.

Theorem 8 *The local marginal polytope can be parametrized by doubly stochastic matrices* $\boldsymbol{\gamma} \in \Delta_{n\times n}$*. With this, the Bethe entropy function* $H_{\mathrm{B}}(\boldsymbol{\gamma})$ *is a concave function of* $\boldsymbol{\gamma}$ *and the Bethe free energy function* $F_{\mathrm{B}}(\boldsymbol{\gamma})$ *is a convex function of* $\boldsymbol{\gamma}$*.*

Proof: Omitted. □

V. Sum-Product-Algorithm-Based Search of the Minimum of the Bethe Free Energy Function

In Section III we have defined the Bethe permanent of a square matrix $\boldsymbol{\theta}$ via the minimum of the Bethe free energy function of the normal factor graph $\mathsf{N}(\boldsymbol{\theta})$. In Section IV we have seen that the Bethe free energy function is a convex function, i.e., it behaves very favorably. This means that we could use any generic optimization algorithm to find the minimum of the Bethe free energy function (see, e.g., [12], [17]) and therefore the Bethe permanent of $\boldsymbol{\theta}$. However, given the special structure of the optimization problem, there is the hope that there are more efficient approaches.

A natural candidate for searching this minimum is the sum-product algorithm (SPA) [14]–[16]. The reason for this is that the theorem by Yedidia, Freeman, and Weiss [8] says that fixed points of the sum-product algorithm correspond to stationary points of the Bethe free energy. Given the convexity of the Bethe free energy function, the only question that remains to be answered is if the SPA always converges to a fixed point. In this section we will show that this is the case, independently of the matrix $\boldsymbol{\theta}$, and (nearly) independently of the chosen initial messages.

Theorem 9 *Consider the SPA for normal factor graph* $\mathsf{N}(\boldsymbol{\theta})$*. With very minor technical conditions on the initial messages, the pseudo-marginals computed by the SPA converge to the pseudo-marginals that minimize the Bethe free energy function of* $\mathsf{N}(\boldsymbol{\theta})$*.*

Proof: Omitted. □

VI. Finite-Graph-Cover Interpretation of the Bethe Permanent

Note that the definition of the permanent in Definition 1 has a "combinatorial flavor." This is in contrast to the definition of the Bethe permanent in Definition 7 that has an "analytical flavor." In this section we show that is possible to represent the Bethe permanent by an expression that has a "combinatorial flavor."

Definition 10 *Consider the non-negative matrix* $\boldsymbol{\theta}$ *of size* $n \times n$*. For every positive integer* M*, we define* Ψ_M *be the set*

$$\Psi_M \triangleq \left\{ \boldsymbol{P} = \left\{ \boldsymbol{P}^{(i,j)} \right\}_{i\in\mathcal{I}, j\in\mathcal{J}} \;\middle|\; \boldsymbol{P}^{(i,j)} \in \mathcal{P}_{M\times M} \right\}.$$

Moreover, for $\boldsymbol{P} \in \Psi_M$ *we define the* $\boldsymbol{P}$*-lifting of* $\boldsymbol{\theta}$ *to be the following* $(nM)\times(nM)$ *matrix*

$$\boldsymbol{\theta}^{\uparrow \boldsymbol{P}} \triangleq \begin{pmatrix} \theta_{1,1}\boldsymbol{P}^{(1,1)} & \cdots & \theta_{1,n}\boldsymbol{P}^{(1,n)} \\ \vdots & & \vdots \\ \theta_{n,1}\boldsymbol{P}^{(n,1)} & \cdots & \theta_{n,n}\boldsymbol{P}^{(n,n)} \end{pmatrix}.$$

Definition 11 *For any positive integer* M*, we define the degree-*M *Bethe permanent of* $\boldsymbol{\theta}$ *to be*

$$\mathrm{perm}_{\mathrm{B},M}(\boldsymbol{\theta}) \triangleq \sqrt[M]{\left\langle \mathrm{perm}\left(\boldsymbol{\theta}^{\uparrow \boldsymbol{P}}\right) \right\rangle_{\boldsymbol{P}\in\Psi_M}},$$

where the angular brackets represent the arithmetic average of $\mathrm{perm}\big(\boldsymbol{\theta}^{\uparrow \boldsymbol{P}}\big)$ *over all* $\boldsymbol{P} \in \Psi_M$*. (Note that the permanent, not the Bethe permanent, appears on the right-hand side of the above expression.)*

Theorem 12 *Consider the non-negative matrix* $\boldsymbol{\theta}$ *of size* $n\times n$*. It holds that*

$$\limsup_{M\to\infty}\ \mathrm{perm}_{\mathrm{B},M}(\boldsymbol{\theta}) = \mathrm{perm}_{\mathrm{B}}(\boldsymbol{\theta}).$$

Proof: Omitted. □

Definition 11, together with Theorem 12 gives the promised "combinatorial characterization" of the Bethe permanent.

VII. Relationship between the Permanent and the Bethe Permanent

Lemma 13 *Let* $\mathbf{1}_{n\times n}$ *be the all-one matrix of size* $n \times n$*. Then*

$$\frac{\mathrm{perm}(\mathbf{1}_{n\times n})}{\mathrm{perm}_{\mathrm{B}}(\mathbf{1}_{n\times n})} = \sqrt{\frac{2\pi n}{\mathrm{e}}} \cdot \big(1 + o(1)\big)$$

Proof: Omitted. □

Theorem 14 *Let* $\boldsymbol{\theta}$ *be an arbitrary non-negative matrix of size* $n \times n$*. Then*

$$\mathrm{perm}_{\mathrm{B}}(\boldsymbol{\theta}) \leqslant \mathrm{perm}(\boldsymbol{\theta}).$$

Proof: Omitted. □

Conjecture 15 *Let* $\boldsymbol{\theta}$ *be an arbitrary non-negative matrix of size* $n \times n$*. We conjecture that*

$$\mathrm{perm}(\boldsymbol{\theta}) \leqslant p(n) \cdot \mathrm{perm}_{\mathrm{B}}(\boldsymbol{\theta})$$

for some suitable function $p(n)$*. In particular, we conjecture that* $p(n) = p \cdot \sqrt{n}$ *for some suitable* p *that is independent of* n*. In fact, we conjecture that* $p(n)$ *is given by the ratio displayed in Lemma 13.*

VIII. Fractional Bethe Entropy

Towards closing the gap between the Bethe permanent and the permanent, the Bethe entropy can be modified following the idea of "fractional Bethe entropy." If the modifications are applied within some suitable limits, the convexity of the (modified) Bethe free energy can be maintained. We omit the details.

Proc. 48th Allerton Conf. on Communications, Control, and Computing,
Allerton House, Monticello, IL, USA, Sep. 29 - Oct. 1, 2010.

IX. Conjectures

It is an interesting challenge to look at theorems involving permaments and to prove that the theorems still hold if the permanents in these theorems are replaced by Bethe permanents. Let us mention two conjectures along these lines.

A. Perm-Pseudo-Codewords

The following conjecture is based on a theorem in [18] involving permanents of submatrices of a parity-check matrix.

Definition 16 *Let $\mathcal{C}$ be a binary linear code described by a parity-check matrix $\boldsymbol{H} \in \mathbb{F}_2^{m \times n}$, $m < n$. For a size-$(m{+}1)$ subset $\mathcal{S}$ of the column index set $\mathcal{I}(\boldsymbol{H})$ we define the Bethe perm-vector based on $\mathcal{S}$ to be the vector $\boldsymbol{\omega} \in \mathbb{Z}^n$ with components*

$$\omega_i \triangleq \begin{cases} \operatorname{perm}_{\mathrm{B}}\left(\boldsymbol{H}_{\mathcal{S}\setminus i}\right) & \text{if } i \in \mathcal{S} \\ 0 & \text{otherwise} \end{cases},$$

where $\boldsymbol{H}_{\mathcal{S}\setminus i}$ is the submatrix of $\boldsymbol{H}$ consisting of all the columns of $\boldsymbol{H}$ whose index is in the set $\mathcal{S} \setminus \{i\}$.

Conjecture 17 *Let $\mathcal{C}$ be a binary linear code described by the parity-check matrix $\boldsymbol{H} \in \mathbb{F}_2^{m \times n}$, $m < n$, and let $\mathcal{S}$ be a size-$(m{+}1)$ subset of $\mathcal{I}(\boldsymbol{H})$. The Bethe perm-vector $\boldsymbol{\omega}$ based on $\mathcal{S}$ is a pseudo-codeword of $\boldsymbol{H}$, i.e.,*

$$\boldsymbol{\omega} \in \mathcal{K}(\boldsymbol{H}). \tag{2}$$

B. Permanent-Based Kernels

Based on a result by Cuturi [19], Huang and Jebara [7] made the following conjecture.

Conjecture 18 ([7]) *Let n be a positive integer and let $\mathcal{X}$ be a set endowed with a kernel κ. Let $X = \{x_1, \ldots, x_n\} \in \mathcal{X}^n$ and $Y = \{y_1, \ldots, y_n\} \in \mathcal{X}^n$. Then*

$$\kappa_{\mathrm{perm}_{\mathrm{B}}} : \ (X, Y) \mapsto \operatorname{perm}_{\mathrm{B}}\left(\left[\kappa(x_i, y_j)\right]_{1 \leqslant i \leqslant n,\, 1 \leqslant j \leqslant n}\right)$$

is a positive definite kernel on $\mathcal{X}^n \times \mathcal{X}^n$.

References

[1] H. Minc, *Permanents*. Reading, MA: Addison-Wesley, 1978.

[2] H. J. Ryser, *Combinatorial Mmathematics (Carus Mathematical Monographs No. 14)*. Mathematical Association of America, 1963.

[3] L. Valiant, "The complexity of computing the permanent," *Theor. Comp. Sc.*, vol. 8, no. 2, pp. 189–201, 1979.

[4] P. Dagum and M. Luby, "Approximating the permanent of graphs with large factors," *Theoretical Computer Science*, vol. 102, no. 2, pp. 283–305, 1992.

[5] 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, July 2004.

[6] M. Chertkov, L. Kroc, and M. Vergassola, "Belief propagation and beyond for particle tracking," *CoRR, available online under* `http://arxiv.org/abs/0806.1199`, June 2008.

[7] B. Huang and T. Jebara, "Approximating the permanent with belief propagation," *CoRR, available online under* `http://arxiv.org/abs/0908.1769`, Aug. 2009.

[8] 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. 2282–2312, July 2005.

[9] Y. Watanabe and M. Chertkov, "Belief propagation and loop calculus for the permanent of a non-negative matrix," *Journal of Physics A: Mathematical and Theoretical*, vol. 43, p. 242002, 2010.

[10] M. Chertkov, L. Kroc, F. Krzakala, M. Vergassola, and L. Zdeborová, "Inference in particle tracking experiments by passing messsages between images," *Proc. Natl. Acad. Sci.*, vol. 107, pp. 7663–8, 2010.

[11] M. Bayati and C. Nair, "A rigorous proof of the cavity method for counting matchings," in *Proc. 44th Allerton Conf. on Communications, Control, and Computing*, Allerton House, Monticello, Illinois, USA, Sep. 27-29 2006.

[12] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[13] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge: Cambridge University Press, 1990, corrected reprint of the 1985 original.

[14] 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.

[15] G. D. Forney, Jr., "Codes on graphs: normal realizations," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 520–548, Feb. 2001.

[16] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Sig. Proc. Mag.*, vol. 21, no. 1, pp. 28–41, Jan. 2004.

[17] D. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.

[18] R. Smarandache and P. O. Vontobel, "Absdet-pseudo-codewords and perm-pseudo-codewords: definitions and properties," in *Proc. IEEE Int. Symp. Information Theory*, Seoul, Korea, June 28–July 3 2009.

[19] M. Cuturi, "Permanents, transportation polytopes and positive definite kernels on histograms," in *Proc. Int. Joint Conf. Artificial Intelligence*, Hyderabad, India, Jan. 6–12 2007.