Double-Edge Factor Graphs:
Definition, Properties, and Examples

Michael X. Cao and Pascal O. Vontobel

Department of Information Engineering

The Chinese University of Hong Kong

{m.x.cao, pascal.vontobel}@ieee.org

Abstract—Some of the most interesting quantities associated with a factor graph are its marginals and its partition sum. For factor graphs *without cycles* and moderate message update complexities, the sum-product algorithm (SPA) can be used to efficiently compute these quantities exactly. Moreover, for various classes of factor graphs *with cycles*, the SPA has been successfully applied to efficiently compute good approximations to these quantities. Note that in the case of factor graphs with cycles, the local functions are usually non-negative real-valued functions.

In this paper we introduce a class of factor graphs, called double-edge factor graphs (DE-FGs), which allow local functions to be complex-valued and only require them, in some suitable sense, to be positive semi-definite. We discuss various properties of the SPA when running it on DE-FGs and we show promising numerical results for various example DE-FGs, some of which have connections to quantum information processing.

I. INTRODUCTION

On the one hand, many classical algorithms like Kalman filtering, the BCJR algorithm, the forward-backward algorithm, *etc.*, can be seen as special cases of the sum-product algorithm (SPA) applied to suitable *cycle-free* factor graphs [1], [2]. On the other hand, the SPA has also been successfully applied to various classes of factor graphs *with cycles*, as is for example witnessed by the SPA-based decoding techniques of low-density parity-check (LDPC) codes, which appear nowadays in various telecommunication standards [1], [2].

For the case of SPA on factor graphs with cycles, there are a few results that hold for large classes of factor graphs (like the result by Yedidia *et al.* [3], which states that fixed points of the SPA correspond to stationary points of the Bethe free energy function) or the graph-cover-based interpretation of the Bethe approximation of the partition sum [4], but in general the results are for special classes of factor graphs like Gaussian graphical models (see, *e.g.*, [5]) or log-supermodular ("attractive") graphical models (see, *e.g.*, [6]). In most of these cases, the focus has been on factor graphs with non-negative real-valued local functions. However, there are applications, in particular in the area of quantum information processing, where one would like to have more general factor graphs. Let us mention some of the approaches that have been pursued:

- One approach replaces scalar-valued local functions by matrix-valued local functions (see, *e.g.*, [7], [8]).
- Another approach keeps scalar-valued local functions, but imposes certain symmetry conditions on the factor graph [9], [10]. (See the discussion and the references in [10] on how the factor graphs therein are related to

tensor networks, *etc.*) The framework in [9], [10] can, for example, be conveniently used for estimating information rates of channels with a classical input and output and a quantum memory [11].

Note that all marginal calculations were done exactly in [9], [10], [11]. This can be achieved, for example, by first merging suitable variables so that the resulting factor graph is cycle free and then to apply the SPA. (Of course, this only gives practical algorithms as long as the alphabet sizes of the merged variables are not too large.)

However, similar to the above-mentioned classes of factor graphs with cycles, it is tempting to also apply the SPA to factor graphs as in [9], [10] with cycles. There are different approaches to accomplish this by suitably reformulating the factor graphs in [9], [10], some reformulations having better complexity properties, some reformulations having better analytical properties. An interesting option in this design space are the double-edge factor graphs (DE-FGs) that we introduce in this paper.[1]

This paper is structured as follows. We define DE-FGs in Section II and then formulate the SPA, along with some of its properties, in Section III. We discuss a variety of examples in Section IV, we point out connections to a recent paper by Mori in Section V, and we conclude the paper in Section VI. Note that throughout this paper, all alphabets are assumed to be finite.

II. DOUBLE-EDGE FACTOR GRAPHS

In this section we define double-edge factor graphs (DE-FGs), more precisely, double-edge normal factor graphs (DE-NFGs). The word "normal" refers to the fact that variables appear as arguments of only one or two local functions.[2]

Example 1. Consider the DE-NFG in Fig. 1, which is a pictorial representation of the factorization

$$g(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{y}) = f_0(y_1, y_2)$$
$$\cdot f_1(x_1, x_2, x_1', x_2'; y_3) \cdot f_2(x_2, x_3, x_5, x_2', x_3', x_5')$$
$$\cdot f_3(x_3, x_4, x_3', x_4'; y_2) \cdot f_4(x_4, x_1, x_4', x_1'; y_1).$$

---

[1] As we will see, the name "double-edge" comes from the fact that pairs of edges (and with that the associated variables) are merged. For example, referring to in Fig. 2 (top), the edge associated with variable x_0 and the edge associated with variable x_0' are merged to a double-edge in Fig. 2 (bottom).

[2] In the same way that any factor graph can be suitably reformulated as a normal factor graph [12], any DE-FG can be suitably reformulated as a DE-NFG. With this, there is no loss in generality in considering only DE-NFGs.
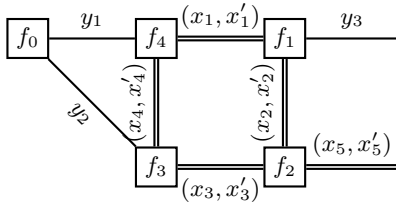
Fig. 1. A double-edge normal factor graph (DE-NFG).

It is called a DE-NFG because some of the edges are double lines that correspond to variables that are paired. (For example, x_1 and x_1' are paired in Fig. 1.) Such paired variables are assumed to have both the same alphabet. Moreover, as detailed below, the local functions have to satisfy some constraints. ▲

Definition 2. Consider the factorization

$$g(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{y}) = \prod_{f \in \mathcal{F}} f(\boldsymbol{x}_{\partial f}, \boldsymbol{x}'_{\partial f}; \boldsymbol{y}_{\delta f})$$

represented by some DE-NFG. We will use the following conventions:

- We call g the global function.
- We call $f \in \mathcal{F}$ the local functions. With some abuse of notation, we will also use f to refer to the corresponding function node in the DE-NFG.
- For every function node $f \in \mathcal{F}$, the variables associated with the incident double-edges are collected in $\boldsymbol{x}_{\partial f}, \boldsymbol{x}'_{\partial f}$.
- For every function node $f \in \mathcal{F}$, the variables associated with the incident single-edges are collected in $\boldsymbol{y}_{\delta f}$.

Most importantly, we require every local function $f \in \mathcal{F}$ to have the following property:

the local function $f(\boldsymbol{x}_{\partial f}, \boldsymbol{x}'_{\partial f}; \boldsymbol{y}_{\delta f})$ is complex-valued and is positive semi-definite (PSD).

The latter property is to be understood as follows: for every $\boldsymbol{y}_{\delta f}$ and every complex-valued function h over the alphabet of $\boldsymbol{x}_{\partial f}$ (and with that also over the alphabet of $\boldsymbol{x}'_{\partial f}$), it holds that

$$\sum_{\boldsymbol{x}_{\partial f}, \boldsymbol{x}'_{\partial f}} \overline{h(\boldsymbol{x}_{\partial f})} \cdot f(\boldsymbol{x}_{\partial f}, \boldsymbol{x}'_{\partial f}; \boldsymbol{y}_{\delta f}) \cdot h(\boldsymbol{x}'_{\partial f}) \geqslant 0. \quad (1)$$

(Here and in the following, over-bar denotes complex conjugation.) Clearly, if a function node f has no incident double edges, then the condition in (1) reduces to the condition that the local function f takes on only non-negative real values. ∎

For proving various properties of DE-NFG, the following observation is very beneficial.

Remark 3. For every local function $f \in \mathcal{F}$ and every $\boldsymbol{y}_{\delta f}$, there are a finite set $\mathcal{K}_{f, \boldsymbol{y}_{\delta f}}$ and some complex-valued functions $b_{f, \boldsymbol{y}_{\delta f}, k}$, $k \in \mathcal{K}_{f, \boldsymbol{y}_{\delta f}}$, over the alphabet of $\boldsymbol{x}_{\partial f}$ such that

$$f(\boldsymbol{x}_{\partial f}, \boldsymbol{x}'_{\partial f}; \boldsymbol{y}_{\delta f}) = \sum_{k \in \mathcal{K}_{f, \boldsymbol{y}_{\delta f}}} \overline{b_{f, \boldsymbol{y}_{\delta f}, k}(\boldsymbol{x}_{\partial f})} \cdot b_{f, \boldsymbol{y}_{\delta f}, k}(\boldsymbol{x}'_{\partial f}) .$$
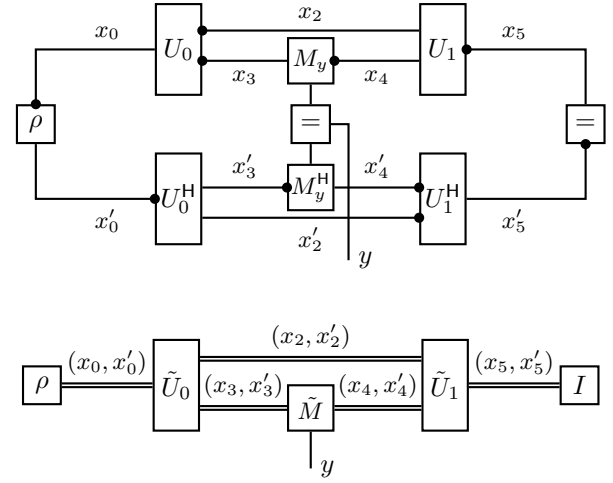

Fig. 2. Top: NFG for Example 5. Bottom: DE-NFG for Example 5.

This follows easily from the eigenvalue decomposition of PSD matrices. ▲

Proposition 4. *The partition sum of a DE-NFG, i.e.,*

$$Z \triangleq \sum_{\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{y}} g(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{y}) ,$$

is a non-negative real number.

Proof. This can be proven with the help of Remark 3. We omit the details because of space limitations. ☐

As already mentioned, one of the main motivations of the present paper are the NFGs in [9], [10]. So let us show how a "typical" NFG in [9], [10] can be formulated as a DE-NFG.

Example 5. Consider the NFG in Fig. 2 (top), which can be used to do probability computations for the following quantum mechanical setup:

- At the beginning, some quantum mechanical system is in some mixed state (represented by the density matrix ρ, which is a PSD matrix).
- The system then evolves unitarily (represented by U_0).
- Afterwards, a sub-system is measured (represented by measurement operators $\{M_y\}_y$).
- Finally the system evolves unitarily (represented by U_1).

(For further details, we refer to [9], [10].) This NFG can be turned into the DE-NFG shown in Fig. 2 (bottom) by suitably merging edges (and with that the associated variables) and by suitably defining the DE-NFG's function nodes. For example, the function node $\tilde{M}$ is defined to be

$$\tilde{M}(x_3, x_4, x_3', x_4'; y) \triangleq M_y(x_3, x_4) \cdot M_y(x_3', x_4') . \quad (2)$$

Clearly, the function $\tilde{M}(x_3, x_4, x_3', x_4'; y)$ satisfies the required PSD constraint. In fact, the expression in (2) is in the form of the decomposition in Remark 3. ▲

One can check that the redrawing procedure in Example 5 can be applied to all relevant NFGs in [9], [10].

III. Sum-Product Algorithm on DE-NFGs and the Bethe Approximation

In this section we define the SPA for DE-NFGs and discuss some of its properties. In particular, we connect it to generalized versions of the Bethe free energy function.

Once a DE-NFG as in Fig. 1 or in Fig. 2 (bottom) has been defined, we simply consider it as a particular type of NFG and apply the SPA in the standard way [1], [2]. Some comments:

- In this paper we only discuss the flooding schedule [1], where all messages are updated at every iteration. Clearly, other update schedules are possible and might be preferable in some cases.
- If desired, message can be rescaled by a positive scalar at every iteration.
- For reasons of simplicity, we discuss only the case where all edges are full edges, *i.e.*, connect two function nodes. (Note that any DE-NFG can be turned into such a DE-NFG by attaching suitable dummy function nodes to half-edges, thereby turning half-edges into full-edges without changing marginals or the partition sum.)

Recall that in the case of NFGs, messages are functions over the alphabet of the variable associated with an edge. Therefore, along a single-edge e between some function nodes f and h, we will have messages $\mu_{e \to f}^{(t)}(y_e)$ and $\mu_{e \to h}^{(t)}(y_e)$ at iteration t. Similarly, along a double-edge between some function nodes f and h, we will have messages $\mu_{e \to f}^{(t)}(x_e, x_e')$ and $\mu_{e \to h}^{(t)}(x_e, x_e')$ at iteration t.

Assumption 6. We make the following assumptions about the initial messages, *i.e.*, about the messages at time $t = 0$:

- Messages along single-edges are positive real-valued functions.
- Messages along double-edges are complex-valued positive definite (PD) functions. ∎

Proposition 7. *Let the messages be initialized as in Assumption 6. Then for every iteration $t \geqslant 1$ it holds that:*

- *Messages along single-edges are non-negative real-valued functions.*
- *Messages along double-edges are complex-valued PSD functions.*

Proof. One approach to prove these statements is based on Remark 3. Another approach is based on Schur's product theorem, which states that the component-wise product of two PSD matrices is a PSD matrix.[3] □

Definition 8. Consider a collection of SPA messages, one for every edge in both directions. Let

$$Z_{\text{Bethe}} \triangleq \prod_{f \in \mathcal{F}} Z_f \Big/ \prod_{e \in \mathcal{E}} Z_e \; ,$$

where $\mathcal{E}$ is the set of all edges, where for every $f \in \mathcal{F}$ we define $Z_f \triangleq \sum_{\boldsymbol{x}_{\partial f}, \boldsymbol{x}_{\partial f}', \boldsymbol{y}_{\delta f}} f(\boldsymbol{x}_{\partial f}, \boldsymbol{x}_{\partial f}'; \boldsymbol{y}_{\delta f}) \cdot$

---

[3] Actually, Schur's product theorem makes the stronger statement that the component-wise product of two PD matrices is a PD matrix.

$\left(\prod_{e \in \partial f} \mu_{e \to f}(x_e, x_e') \right) \cdot \left(\prod_{e \in \delta f} \mu_{e \to f}(y_e) \right)$, where for every single-edge $e \in \mathcal{E}$ between function nodes f and h we define $Z_e \triangleq \sum_{y_e} \mu_{e \to f}(y_e) \cdot \mu_{e \to h}(y_e)$, and where for every double-edge $e \in \mathcal{E}$ between function nodes f and h we define $Z_e \triangleq \sum_{x_e, x_e'} \mu_{e \to f}(x_e, x_e') \cdot \mu_{e \to h}(x_e, x_e')$. ∎

Proposition 9. *The function Z_{Bethe} in Definition 8 has the following properties:*

- *Assume that the messages have the properties in Proposition 7 and assume that Z_{Bethe} is well-defined, i.e., $Z_e \neq 0$ for all $e \in \mathcal{E}$. Then Z_{Bethe} is a non-negative real number.*
- *Fixed points of the SPA correspond to stationary points of the function Z_{Bethe}. (This generalizes a theorem by Yedidia et al. [3].)*

Proof. Omitted due to space limitations. □

Evaluating Z_{Bethe} in Definition 8 at a fixed point of the SPA results in the Bethe approximation of the partition sum of the DE-NFG.

One can also generalize the Bethe free energy function F_{Bethe} from [3], where F_{Bethe} is a function over a suitable generalization of the local marginal polytope. While a statement (analogous to a statement in [3]) that fixed points of the SPA correspond to stationary points of F_{Bethe} can be made, evaluating Z_{Bethe} based on F_{Bethe} is trickier because of the multi-valuedness of the complex logarithm.

IV. Examples

In this section we discuss various examples of DE-NFGs. In particular, we compare the obtained Bethe approximation of the partition sum with the true partition sum. (The NFGs in this section have modest sizes so that the true partition function can be computed efficiently.) Moreover, for the first example, we can also make some analytical statements.

Example 10. Let n be some integer larger than one. Consider a DE-NFG whose topology is an n-cycle and where all variables take on values in the same finite alphabet $\mathcal{X}$. (Fig. 3 shows such a DE-NFG for $n = 4$.) Let $\boldsymbol{F}$ be a complex-valued PD matrix of size $|\mathcal{X}|^2 \times |\mathcal{X}|^2$ with entries $F(x_0, x_1; x_0', x_1')$. For $i \in [n] \triangleq \{0, 1, \ldots, n-1\}$, we define the local function f_i to be $f_i(x_i, x_{i+1}; x_i', x_{i+1}') \triangleq F(x_i, x_{i+1}; x_i', x_{i+1}')$. (All indices are modulo n.)

In order to proceed, it is convenient to define the complex-value matrix $\boldsymbol{B}$ of size $|\mathcal{X}|^2 \times |\mathcal{X}|^2$ with entries $B(x_0, x_0'; x_1, x_1') \triangleq F(x_0, x_1; x_0'; x_1')$ and to define $\tilde{x}_i \triangleq (x_i, x_i')$, $i \in [n]$.

Let $\mu_{f_i \to f_{i+1}}^{(t)}(\tilde{x}_{i+1})$ be the SPA message along the double edge from f_i to f_{i+1} at time index t. Similarly, let $\mu_{f_{i+1} \to f_i}^{(t)}(\tilde{x}_{i+1})$ be the SPA message along the double edge from f_{i+1} to f_i at time index t. Clearly,

$$\mu_{f_i \to f_{i+1}}^{(t)}(\tilde{x}_{i+1}) \propto \sum_{\tilde{x}_{i-1}} \mu_{f_{i-1} \to f_i}^{(t-1)}(\tilde{x}_{i-1}) \cdot B(\tilde{x}_{i-1}, \tilde{x}_i) \; , \quad (3)$$

$$\mu_{f_i \to f_{i-1}}^{(t)}(\tilde{x}_i) \propto \sum_{\tilde{x}_{i+1}} B(\tilde{x}_i, \tilde{x}_{i+1}) \cdot \mu_{f_{i+1} \to f_i}^{(t-1)}(\tilde{x}_{i+1}) \; . \quad (4)$$
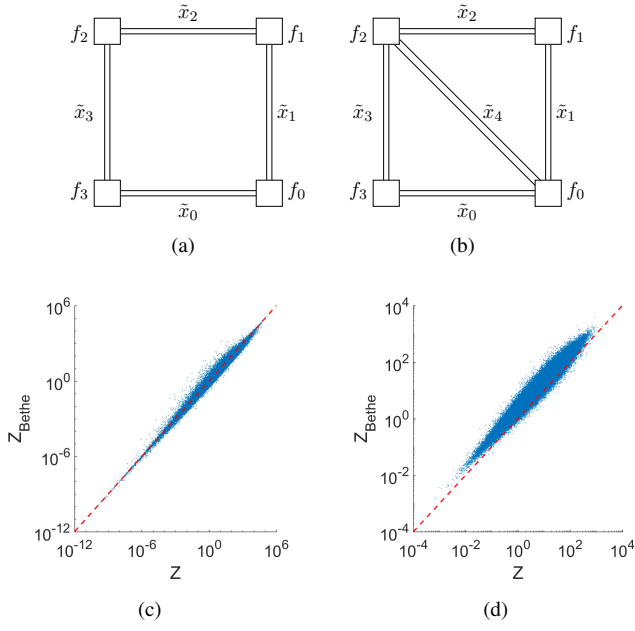
Fig. 3. DE-NFGs used in Examples 10 and 11.

Fig. 4. DE-NFGs used in Example 12.

For $i \in [n]$, we assume the following initializations $\mu_{f_i \to f_{i+1}}^{(0)}(\tilde{x}_{i+1}) \triangleq \delta(x_{i+1}, x'_{i+1})$ and $\mu_{f_{i+1} \to f_i}^{(0)}(\tilde{x}_{i+1}) \triangleq \delta(x_{i+1}, x'_{i+1})$, where δ is the Kronecker-delta function.

Because of the properties of the matrix $\boldsymbol{B}$ that are induced by the properties of the matrix $\boldsymbol{F}$, the SPA message update rules in (3)–(4) represent so-called completely positive maps (see, *e.g.*, [13]). (For this statement we ignore the rescaling factors.) Using generalizations of Perron–Frobenius theory (see [14], [15]), one can make the following statements:

- For every $i \in [n]$, the message $\mu_{f_i \to f_{i+1}}^{(t)}(\tilde{x}_{i+1})$ converges to a PD matrix as $t \to \infty$.
- For every $i \in [n]$, the message $\mu_{f_{i+1} \to f_i}^{(t)}(\tilde{x}_{i+1})$ converges to a PD matrix as $t \to \infty$.
- The eigenvalue of the matrix $\boldsymbol{B}$ with maximum absolute value is a real number and is unique. Let us call it λ_0.
- The Bethe approximation of the partition sum is

$$Z_{\text{Bethe}} = \lambda_0^n . \tag{5}$$

Compare this result with the partition sum, which is

$$Z = \sum_{j=0}^{|\mathcal{X}|^2-1} \lambda_j^n = \lambda_0^n \cdot \left(1 + \sum_{j=1}^{|\mathcal{X}|^2-1} \left(\frac{\lambda_j}{\lambda_0} \right)^n \right) , \tag{6}$$

where $\lambda_0, \ldots, \lambda_{|\mathcal{X}|^2-1}$ are the eigenvalues of $\boldsymbol{B}$. We see that the smaller the ratios $\left(\frac{\lambda_j}{\lambda_0} \right)^n$, $j = 1, \ldots, |\mathcal{X}|^2-1$, are, the better the Bethe approximation is.

For $n = 4$ and $|\mathcal{X}| = 2$, Fig. 3(c) shows the obtained Z and Z_{Bethe} values for 10^6 experiments based on randomly generating matrices $\boldsymbol{F} \triangleq \boldsymbol{U} \cdot \boldsymbol{D} \cdot \boldsymbol{U}^{\mathsf{H}}$, which are based on randomly generating unitary matrices $\boldsymbol{U}$ and diagonal matrices $\boldsymbol{D}$, where the diagonal entries of $\boldsymbol{D}$ are sampled i.i.d. from a standard χ^2 distribution with one degree of freedom. We see that very often the ratio Z_{Bethe}/Z is rather close to 1. ▲
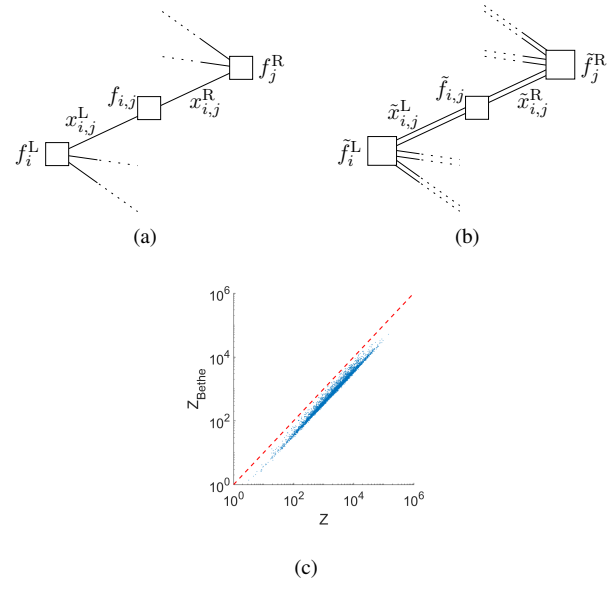
Example 11. Consider now the DE-NFG in Fig. 3(b). For $|\mathcal{X}| = 2$, Fig. 3(d) shows the obtained Z and Z_{Bethe} values for 10^6 experiments based on randomly generating local functions. In contrast to Example 10, where for every instantiation all local function were the same, here for every instantiation all local function are generated independently. We observe the ratio Z_{Bethe}/Z is reasonably close to 1, but typically larger than 1. ▲

Example 12. Let $\boldsymbol{\theta}$ be a complex-valued matrix of size $n \times n$ with entries $\theta_{i,j}$. The permanent [16] of $\boldsymbol{\theta}$ is defined to be $\text{perm}(\boldsymbol{\theta}) \triangleq \sum_{\sigma} \prod_{i=1}^n \theta_{i,\sigma(i)}$, where the summation is over all $n!$ permutations of the set $[n] \triangleq \{1, \ldots, n\}$. Ryser's algorithm, one of the most efficient algorithms for exactly computing $\text{perm}(\boldsymbol{\theta})$ for general matrices θ, requires $\Theta(n \cdot 2^n)$ arithmetic operations [17], and so the exact computation of permanent is intractable, even for moderate values of n. Note that even the computation of the permanent of matrices that contain only zeros and ones is #P-complete [18].

One can formulate an NFG whose partition sum equals $\text{perm}(\boldsymbol{\theta})$, see, *e.g.*, Fig. 1 in [19]. That NFG is a complete bipartite graph with n function nodes on the left and n function nodes on the right. Here, Fig. 4(a), shows a slightly modified version of that NFG. All variables take values in the set $\mathcal{X} = \{0, 1\}$. Moreover, for all $i \in [n]$, the function f_i^{L} is defined to be

$$f_i^{\text{L}}\big(\{x_{i,j}^{\text{L}}\}_{j \in [n]}\big) \triangleq \begin{cases} 1 & \text{exactly one of } \{x_{i,j}^{\text{L}}\}_{j \in [n]} \text{ equals } 1 \\ 0 & \text{otherwise} \end{cases} ;$$

for all $j \in [n]$, the function f_j^{R} is defined analogously; and for all $(i,j) \in [n]^2$, the function $f_{i,j}$ is defined to be

$$f_{i,j}\big(x_{i,j}^{\text{L}}, x_{i,j}^{\text{R}}\big) \triangleq \delta\big(x_{i,j}^{\text{L}}, x_{i,j}^{\text{R}}\big) \cdot \begin{cases} \theta_{i,j} & \text{if } x_{i,j}^{\text{L}} = 1 \\ 1 & \text{if } x_{i,j}^{\text{L}} = 0 \end{cases} .$$

In this example, we consider the following, rather natural generalization to the DE-NFG in Fig. 4(b), where we will use the short-hand $\tilde{x}_{i,j}^{\mathrm{L}}$ for $\big(x_{i,j}^{\mathrm{L}}, \tilde{x}_{i,j}^{\mathrm{L}'}\big)$, *etc.* Assume that for $(i,j) \in [n]^2$, $\tilde{\theta}_{i,j}$ is a complex-valued PSD matrix of size 2×2. With this, for $i \in [n]$, the function $\tilde{f}_i^{\mathrm{L}}$ is defined to be

$$\tilde{f}_i^{\mathrm{L}}\big(\{\tilde{x}_{i,j}^{\mathrm{L}}\}_{j\in[n]}\big) \triangleq f_i^{\mathrm{L}}\big(\{x_{i,j}^{\mathrm{L}}\}_{j\in[n]}\big) \cdot f_i^{\mathrm{L}}\big(\{x_{i,j}^{\mathrm{L}'}\}_{j\in[n]}\big) \ ;$$

for all $j \in [n]$, the function $\tilde{f}_j^{\mathrm{R}}$ is defined analogously; and for all $(i,j) \in [n]^2$, the function $\tilde{f}_{i,j}$ is defined to be

$$\tilde{f}_{i,j}\big(\tilde{x}_{i,j}^{\mathrm{L}}, \tilde{x}_{i,j}^{\mathrm{R}}\big) \triangleq \delta\big(x_{i,j}^{\mathrm{L}}, x_{i,j}^{\mathrm{R}}\big) \cdot \delta\big(x_{i,j}^{\mathrm{L}'}, x_{i,j}^{\mathrm{R}'}\big) \cdot \tilde{\theta}_{i,j}\big(x_{i,j}^{\mathrm{L}}, x_{i,j}^{\mathrm{L}'}\big) \ .$$

(One can easily verify that these local function define indeed a DE-NFG.) Finally, let $\tilde{Z}$ be the partition sum of this DE-NFG.

This DE-NFG definition has the following two important special cases:

- If $\tilde{\theta}_{i,j} = \big(\begin{smallmatrix} 1 & 0 \\ 0 & \theta_{i,j} \end{smallmatrix}\big)$ for all $(i,j) \in [n]^2$, then $\tilde{Z} = \mathrm{perm}(\boldsymbol{\theta})$.
- If $\tilde{\theta}_{i,j} = \big(\begin{smallmatrix} 1 \\ \theta_{i,j} \end{smallmatrix}\big) \cdot \big(\begin{smallmatrix} 1 & \overline{\theta_{i,j}} \end{smallmatrix}\big) = \big(\begin{smallmatrix} 1 & \overline{\theta_{i,j}} \\ \theta_{i,j} & |\theta_{i,j}|^2 \end{smallmatrix}\big)$ for all $(i,j) \in [n]^2$, then $\tilde{Z} = \mathrm{perm}(\boldsymbol{\theta}) \cdot \mathrm{perm}(\overline{\boldsymbol{\theta}}) = \big|\mathrm{perm}(\boldsymbol{\theta})\big|^2$, where $\overline{\boldsymbol{\theta}}$ denotes the matrix whose entries are the complex-conjugate values of the entries of $\boldsymbol{\theta}$. (Note that such partition sums are of interest in quantum information processing [20], where $\boldsymbol{\theta}$ are certain types of square matrices over the complex numbers. We refer to [20] for details.)

In our experiments, we considered the following setup. Namely, for every $(i,j) \in [n]^2$, we independently generate $\tilde{\theta}_{i,j}$ as follows: $\tilde{\theta}_{i,j}(0,0) \triangleq 1$; $\tilde{\theta}_{i,j}(1,0)$ is picked uniformly from the unit circle in the complex plane; $\tilde{\theta}_{i,j}(0,1) \triangleq \overline{\tilde{\theta}_{i,j}(1,0)}$; $\tilde{\theta}_{i,j}(1,1)$ is picked uniformly (and independently of the other entries) from the real line interval $[1.10, \ 11.10]$. Fig. 4(c) shows the obtained Z and Z_{Bethe} values for 5000 experiments for the case $n = 5$. We observe that the ratio Z_{Bethe}/Z is concentrated around a value smaller than 1. $\blacktriangle$

V. Connections to a Paper by Mori

Finally, let us point out that there are strong connections of DE-NFGs to the setup in Section V of a recent paper by Mori [21]. Assume to have a bipartite DE-NFG. (Such a DE-NFG can always be obtained by suitably inserting dummy function nodes.) Then the partition sum can be written as some inner product between, on the one hand, the tensor product of the local functions corresponding to first class of function nodes of the bipartite DE-NFG, and, on the other hand, the tensor product of the local functions corresponding to the second class of function nodes of the bipartite DE-NFG. Once this connection is observed, one can translate Mori's results (like loop calculus expansions) to DE-NFGs.

VI. Conclusion

In this paper we have defined DE-NFGs and studied some of their properties. In particular, we have shown some promising numerical studies of the Bethe approximation to the partition sum. Many open questions remain. For example, can some of the results in [19] be generalized to the setup in Example 12?

Or, as in the context of computing the pattern maximum likelihood estimate, which can be formulated as optimizing the parameters of some graphical model toward maximizing the partition function, and where the Bethe partition sum was beneficially used as a surrogate function [22], can the Bethe partition sum of a DE-NFG serve as a suitable surrogate function in some partition function optimization problem?

References

[1] 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.

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

[3] 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, Jul. 2005.

[4] P. O. Vontobel, "Counting in graph covers: a combinatorial characterization of the Bethe entropy function," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 6018–6048, Sep. 2013.

[5] D. M. Malioutov, J. K. Johnson, and A. S. Willsky, "Walk-sums and belief propagation in Gaussian graphical models," *J. Mach. Learn. Res.*, vol. 7, pp. 2031–2064, Dec. 2006.

[6] N. Ruozzi, "The Bethe partition function of log-supermodular graphical models," in *Proc. Neural Inf. Proc. Sys. Conf.*, Lake Tahoe, NV, USA, Dec. 3–6 2012.

[7] M. S. Leifer and D. Poulin, "Quantum graphical models and belief propagation," *Ann. Phys.*, vol. 323, pp. 1899–1946, 2008.

[8] M. X. Cao and P. O. Vontobel, "Quantum factor graphs: closing-the-box operation and variational approaches," in *Proc. Int. Symp. Inf. Theory and its Appl.*, Monterey, CA, USA, Oct. 30–Nov. 2 2016, pp. 651–655.

[9] H.-A. Loeliger and P. O. Vontobel, "A factor-graph representation of probabilities in quantum mechanics," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 1–6 2012, pp. 656–660.

[10] ——, "Factor graphs for quantum probabilities," *accepted for IEEE Trans. Inf. Theory, available online under* http://arxiv.org/abs/1508.00689, 2017.

[11] M. X. Cao and P. O. Vontobel, "Estimating the information rate of a channel with classical input and output and a quantum state," *to appear in Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, extended version available under https://arxiv.org/abs/1705.01041.

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

[13] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge, UK: Cambridge University Press, 2000.

[14] D. E. Evans and R. Hoegh-Krohn, "Spectral properties of positive maps on C^* algebras," *J. London Math. Soc.*, vol. 17, no. 2, pp. 345–355, 1978.

[15] R. Schrader, "Perron–Frobenius theory for positive maps on trace ideals," in *Mathematical Physics in Mathematics and Physics: Quantum and Operator Algebraic Aspects*, ser. Fields Inst. Commun., R. Longo, Ed. Providence, RI, USA: Amer. Math. Soc., 2001, available online at https://arxiv.org/abs/math-ph/0007020.

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

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

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

[19] P. O. Vontobel, "The Bethe permanent of a non-negative matrix," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1866–1901, Mar. 2013.

[20] S. Aaronson and A. Arkhipov, "The computational complexity of linear optics," *Theory of Computation*, vol. 9, no. 4, pp. 143–252, 2013.

[21] R. Mori, "Holographic transformation, belief propagation and loop calculus for generalized probabilistic theories," in *Proc. IEEE Int. Symp. Inf. Theory*, Hong Kong, China, June 14–19 2015, pp. 1099–1103.

[22] P. O. Vontobel, "The Bethe approximation of the pattern maximum likelihood distribution," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 1–6 2012, pp. 2012–2016.