On the Relationship Between the Global Minimum of the Bethe Free Energy Function of a Factor Graph and Sum-Product Algorithm Fixed Points

Yuwen Huang and Pascal O. Vontobel
Department of Information Engineering
The Chinese University of Hong Kong
hy018@ie.cuhk.edu.hk, pascal.vontobel@ieee.org

Abstract

The sum-product algorithm (SPA) is a popular algorithm for efficiently approximating the marginals and the partition function of a factor graph. Some key results for this algorithm were established by Yedidia *et al.*, who proved that, roughly speaking, fixed points of the SPA correspond to stationary points of the Bethe free energy function. However, some of their results were only for factor graphs where the local functions take on strictly positive values. They also conjectured that similar results hold for factor graphs where the local functions take on non-negative values. In this paper we make progress toward resolving this conjecture. In particular, we present examples where the results of Yedidia *et al.* generalize and examples where their results do not generalize. Finally, we present a general framework for analyzing fixed-points of the SPA based on a suitable dualization of the Bethe free energy function.

I. Introduction

In this paper, we consider standard factor graphs (S-FGs), which are factor graphs [1]–[3] where all local functions take on non-negatives values. S-FGs are used in a wide variety of disciplines, including communications (see, e.g., [4]), statistical mechanics (see, e.g., [5]), and coding theory (see, e.g., [6]). Inference problems involving probabilistic models in these areas are often formulated as computing the marginal probabilities of some subsets of the variables in some S-FG and/or computing the partition function of some S-FG.

The sum-product algorithm (SPA), also known as loopy belief propagation (LBP), is a practical and powerful way to approximately compute the marginals and the partition function of an S-FG. In the case of cycle-free S-FGs, the SPA provides the exact marginals and partition function. In the case of S-FG with cycles, the SPA often gives surprisingly good approximations of the marginals and the partition function. This is part of the reason why SPA decoding of low-density parity-check (LDPC) codes appears in the 5G telecommunications standard. Nevertheless, there are also S-FGs where the SPA provides a poor approximation of the marginals or the SPA even fails to converge [7], [8].

The seminal paper [9] by Yedidia *et al.* related the SPA fixed points to the Bethe free energy function, which is a function of beliefs. In particular, they proved the following statements.

1) Every interior stationary point of the associated Bethe free energy function corresponds to an SPA fixed point [9, Th. 2].

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 14208319).

2) For S-FGs with positive-valued local functions only, every stationary point of the associated Bethe free energy function corresponds to an SPA fixed point [9, Th. 3].

In a follow-up work, Heskes [8] presented the following results.

1) The stable fixed points of the SPA are minima of the Bethe free energy function [8, Section 4].
2) The minimization problem of the Bethe free energy function can be transformed into a minimax optimization problem [8, Section 4].
3) The fixed point of the algorithm in [8, Algorithm 2] corresponds to a local optimal solution to the transformed minimax problem.

In general, there is significant evidence that the SPA is well behaved also for S-FGs where not all local functions take on only strictly positive values. In particular, the papers [10] (on S-FGs whose partition function equals the permanent of a non-negative matrix) and [11] (on $(2,k)$-regular LDPC codes) established that the SPA finds the global minimum of the Bethe free energy function, even if that global minimum happens to be on the boundary of the Bethe free energy function domain.

There are various other results in the literature about fixed points of the SPA and the minimum of the Bethe free energy. For example, in [12] the authors tried to find all fixed points using the numerical polynomial-homotopy-continuation (NPHC) method, and in [13], the authors analyzed the SPA on patch potential models and obtained interesting insights about the SPA's properties. The authors in [14], [15] studied the progress towards the minimum of the Bethe free energy function by introducing a pseudo-dual function of the Bethe free energy function.

A. *Outline of Results*

In this paper we investigate whether the global minimum of the Bethe free energy function of an S-FG corresponds to an SPA fixed point. Here are the main results of this paper.

1) We present the primal and dual formulations of the Bethe partition function, which are optimization problems whose optimal values are equal. The primal formulation based on [9] is related to the minimization of the Bethe free energy function. The primal formulation has the following properties.
    a) In this minimization problem, the feasible set of the beliefs, is a convex and compact set, which means that the locations of the optimal value are attainable in this set.
    b) The Bethe free energy function is a function of the beliefs associated with the function nodes and the edges in the considered S-FG.
    c) The Bethe free energy function is neither convex nor concave for general S-FG. (For some special S-FGs, e.g., a single-cycle S-FG [16, Corollary 2], the associated Bethe free energy functions are convex.)

    The dual formulation is based on [8], which is a maximin optimization problem where the minimization is taken over a part of the variables and the maximization is taken over the remaining variables. This optimization problem has the following properties.
    a) The feasible set for the variables that are related to the minimization, is the set of real numbers, i.e., an open set, and the locations of the optimal value are allowed to be outside the feasible set, i.e, some of the variables go to infinity when they achieve the optimal value.
    b) The feasible set of the variables related to the maximization, is a compact set.
    c) This maximin optimization problem contains the variables that are associated only with the edges in the considered S-FG.

   d) The associated objective function is convex with respect to (w.r.t.) to the variables that are related to the minimization. (For some special S-FGs, e.g., a single-cycle S-FG [16], the objective functions are concave w.r.t. the variables that are related to the maximization.)

2) To solve the above-mentioned maximin optimization problem that is related to the dual formulation, we propose two algorithms in Algorithms 1 and 2, where Algorithm 1 is based on [8, Algorithm 2], and Algorithm 2 is equivalent to the SPA. We make a comparison between these two algorithms.
   a) Algorithm 1 is a double-loop algorithm while Algorithm 2 is simpler and contains a single loop only.
   b) Algorithms 1 and 2 have the same set of fixed points. However, they have different sets of stable fixed points. For some special S-FGs, Algorithm 1 converges to the location of the optimal value of the maximin optimization problem, while Algorithm 2, i.e., the SPA, fails to converge. For details, see [8, Figure 2].
   c) Both Algorithms 1 and 2 try to find the stationary points of the objective function in the maximin optimization problem.
   d) In particular, the inner loop of Algorithm 1 attempts to find the stationary point of the objective function w.r.t. the variables in the minimization problem. Because the objective function is convex w.r.t. these variables as previously mentioned, the inner loop also finds the locations of the optimal value of the minimization problem in the maximin optimization problem.
3) In order to appreciate the main results for the general S-FG, we first analyze the behavior of the SPA on some simple and interesting S-FGs in Figs. 3, 5, and 6. In particular, for the S-FG in Fig. 3, we cannot evaluate the beliefs at the SPA fixed point. For Figs 5 and 6, we show that the locations of the optimal values of the primal and dual formulations of the Bethe partition function are related to the SPA fixed points.
4) For general S-FG, we show that there exists a sequence of messages such that the beliefs defined based on the messages converge to the locations of the global minimum of the Bethe free energy function. Also the messages converge to one of the SPA fixed points of a modified S-FG that has the same minimum of the Bethe free function as the original S-FG.

We also make a comparison between the dualization in [9], [14], [15] and the dualization in [8], which is also the dualization considered in this paper.

1) The dualization proposed in [9], [14], [15] works for the S-FG such that at least one of the locations of the associated Bethe free energy function's global minimum is in the interior of the feasible set defined in the primal formulation.
2) The dualization proposed in [8] works for any S-FG.
3) The structure of the objective function in the dual formulation proposed in [8] is similar to the structure of the pseudo partition function evaluated at an SPA fixed point, which indicates that there is a relationship between the SPA fixed point and the locations of the optimal value for the dual formulation.

In the following, we will use, without essential loss of generality, normal factor graphs (NFGs), i.e., factor graphs where variables are associated with edges [2], [3].

The rest of this paper is structured as follows. Section II reviews the basics of S-FGs and the associated SPA. Section III presents the primal and the dual formulations of the Bethe partition function. Sections IV, V, and V study the primal and the dual formulations of Bethe partition function as well as the behavior the associated SPA for S-FGs in Figs. 3, 5, and 6, respectively. Section VII considers general S-FGs and relates the locations of the global minimum of the Bethe free energy function to an SPA fixed point.

B. Basic Notations and Definitions

The sets $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, and $\mathbb{R}_{>0}$ are defined to be the field of real numbers, the set of nonnegative real numbers, and the set of positive real numbers. If not mentioned otherwise, all variable alphabets are assumed to be finite. Square brackets are used in two different

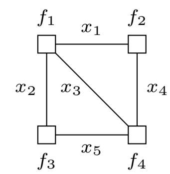


Fig. 1: NFG N in Example 1.

ways. Namely, for any $L \in \mathbb{Z}_{>0}$, the function $[L]$ is defined to be the set $[L] \triangleq \{1, \ldots, L\}$ with cardinality L and for any statement S, by the Iverson's convention, the function $[S]$ is defined to be $[S] \triangleq 1$ if S is true and $[S] \triangleq 0$ otherwise.

II. The Standard Normal Factor Graph (S-NFG)

Factor graph is a convenient way to depict the factorization of a multivariate function [1]. Also many operations, e.g., taking the summations and multiplications for multivariate functions can be visualized by factor graphs. In this paper, we will use, without essential loss of generality, standard normal factor graphs (S-NFGs), i.e., S-FGs where variables are associated with edges [2], [3].

The key aspects of an S-NFG are best explained with the help of an example.

Example 1. *Let us consider the function*

$$g(x_1, \ldots, x_5) \triangleq f_1(x_1, x_2, x_3) \cdot f_2(x_1, x_4) \cdot f_3(x_2, x_5) \cdot f_4(x_3, x_4, x_5).$$

In particular, the function g, the so-called global function, is the product of the so-called local functions $f_1, \ldots, f_4$. The factorization of g can be visualized via the S-NFG N in Fig. 1, where N consists of five (full) edges with associated variables $x_1, \ldots, x_5$, and four function nodes $f_1, \ldots, f_4$.

In general, edge e is incident on function node f if and only if x_e is a argument of the associated local function f.

An edge that connects to two function nodes is called a full edge, whereas an edge that is connected to only one function node is called a half edge. For simplicity, we consider S-NFGs with only full edges, due to the fact that S-NFGs with half edges can be turned into the considered S-NFGs by adding dummy 1-valued function nodes to the half edges without changing any marginal or the partition function.

Definition 2. *An S-NFG $\mathsf{N}(\mathcal{F}(\mathsf{N}), \mathcal{E}(\mathsf{N}), \mathcal{X}(\mathsf{N}))$ consists of the following objects:*

1) *The graph $(\mathcal{F}(\mathsf{N}), \mathcal{E}(\mathsf{N}))$ with vertex set $\mathcal{F}(\mathsf{N})$ and edge set $\mathcal{E}(\mathsf{N})$, where $\mathcal{F}(\mathsf{N})$ is also known as the set of function nodes. (Every edge $e \in \mathcal{E}(\mathsf{N})$ will be assumed to be a full edge connecting two function nodes.) Suppose that the numbers of function nodes and edges are F and E. The order of elements in the function node set and edge set is fixed*

$$\mathcal{F}(\mathsf{N}) = \{f_1, \quad f_2, \quad \ldots \quad, f_F\}, \qquad \mathcal{E}(\mathsf{N}) = [E].$$

2) *The alphabet $\mathcal{X}(\mathsf{N}) \triangleq \prod_{e \in \mathcal{E}(\mathsf{N})} \mathcal{X}_e$, where $\mathcal{X}_e$ is the alphabet of the variable x_e associated with edge $e \in \mathcal{E}$.*

In the following, if there is no ambiguity, we simply use $\mathcal{F}$, $\mathcal{E}$, and $\mathcal{X}$ for $\mathcal{F}(\mathsf{N})$, $\mathcal{E}(\mathsf{N})$, and $\mathcal{X}(\mathsf{N})$, respectively.

Definition 3. *Given $\mathsf{N}(\mathcal{F}, \mathcal{E}, \mathcal{X})$, we make the following definitions:*

1) *For every function node $f \in \mathcal{F}$, the set ∂f is defined to be the set of edges incident on f.*

2) *For every edge* $e = (f_i, f_j) \in \mathcal{E}$ *such that* $i < j$,[1] *the pair* (f_i, f_j) *is defined to be the pair of function nodes that are connected to edge* e.
3) *An assignment* $\boldsymbol{x} \triangleq (x_e)_{e\in\mathcal{E}} \in \mathcal{X}$ *is called a configuration of the S-NFG. For each* $f \in \mathcal{F}$, *a configuration* $\boldsymbol{x} \in \mathcal{X}$ *induces the vector* $\boldsymbol{x}_f \triangleq (x_e)_{e\in\partial f}$.
4) *In general, for every* $f \in \mathcal{F}$, *the local function associated with* f *is, with some slight abuse of notation, also called* f.[2] *Here, the local function* f *can be an arbitrary mapping from* $\prod_{e\in\partial f} \mathcal{X}_e$ *to* $\mathbb{R}_{\geq 0}$.
5) *For every* $f \in \mathcal{F}$, *we define the alphabet* $\mathcal{X}_f$ *to be* $\mathcal{X}_f \triangleq \left\{\boldsymbol{x}_f \in \prod_{e\in\partial f} \mathcal{X}_e \;\middle|\; f(\boldsymbol{x}_f) \neq 0\right\}$.
6) *The global function* g *is defined to be the mapping* $g:\ \mathcal{X} \to \mathbb{R}_{\geq 0},\ \boldsymbol{x} \mapsto \prod_{f\in\mathcal{F}} f(\boldsymbol{x}_f)$.
7) *The partition function is defined to be* $Z(\mathsf{N}) \triangleq \sum_{\boldsymbol{x}\in\mathcal{X}} g(\boldsymbol{x})$.[3] *Clearly, the partition function satisfies* $Z(\mathsf{N}) \in \mathbb{R}_{\geq 0}$.
8) *If* $Z(\mathsf{N}) > 0$, *the probability mass function (PMF) induced on* N *is defined to be the function*
$$p(\boldsymbol{x}) \triangleq \frac{g(\boldsymbol{x})}{Z(\mathsf{N})}.$$
9) *Let* $\mathcal{I}$ *be a subset of* $\mathcal{E}(\mathsf{N})$ *and let* $\mathcal{I}^{\mathrm{c}} \triangleq \mathcal{E}(\mathsf{N}) \setminus \mathcal{I}$ *be its complement. The marginal* $p_{\mathcal{I}}(\boldsymbol{x}_{\mathcal{I}})$ *is defined to be*
$$p_{\mathcal{I}}(\boldsymbol{x}_{\mathcal{I}}) \triangleq \sum_{\boldsymbol{x}_{\mathcal{I}^{\mathrm{c}}}} p(\boldsymbol{x}), \qquad \boldsymbol{x}_{\mathcal{I}} \in \mathcal{X}_e^{|\mathcal{I}|},$$
where
$$\boldsymbol{x}_{\mathcal{I}} \triangleq (x_e)_{e\in\mathcal{I}} \in \prod_{e\in\mathcal{I}} \mathcal{X}_e, \qquad \boldsymbol{x}_{\mathcal{I}^{\mathrm{c}}} \triangleq (x_e)_{e\in\mathcal{I}^{\mathrm{c}}} \in \prod_{e\in\mathcal{I}^{\mathrm{c}}} \mathcal{X}_e.$$
If $\mathcal{I} = \{e\}$, *then we have*
$$p_{\{e\}}(x_e) \triangleq \sum_{\boldsymbol{z}:\, z_e = x_e} p(\boldsymbol{z}).$$

For simplicity, when there is no ambiguity, we use the shorthands $\sum_e$, $\sum_f$, $\sum_{x_e}$, $\sum_{\boldsymbol{x}_f}$, $\sum_{\boldsymbol{z}_f}$, $\{\cdot\}_{x_e}$, and $(\cdot)_{x_e}$ for $\sum_{e\in\mathcal{E}}$, $\sum_{f\in\mathcal{F}}$, $\sum_{x_e\in\mathcal{X}_e}$, $\sum_{\boldsymbol{x}_f\in\mathcal{X}_f}$, $\sum_{\boldsymbol{z}_f\in\mathcal{X}_f}$, $\{\cdot\}_{x_e\in\mathcal{X}_e}$ and $(\cdot)_{x_e\in\mathcal{X}_e}$. respectively.

In this paper, we make the following general assumption about S-NFGs.

Assumption 4. *In this paper, for an S-NFG* N *we assume that*
$$\exists \boldsymbol{x} \in \mathcal{X} \text{ such that } g(\boldsymbol{x}) > 0,$$
which is equivalent to assuming that $Z(\mathsf{N}) > 0$.

As mentioned in the introduction section, for an S-NFG N, the SPA often gives a surprisingly good approximation of the partition function $Z(\mathsf{N})$ and the marginal $p_{\{e\}}$ for edge $e \in \mathcal{E}(\mathsf{N})$. Here we present the SPA by providing the technical details only. For the motivation behind the SPA, we refer to [1]–[3].

Definition 5. *Given some S-NFG* N. *The SPA [1]–[3] is an iterative algorithm where the messages, which are functions associated with edges, are sent along edges at each iteration. In particular, at each iteration, two messages are sent along each edge, one in both directions.) The SPA consists of the following steps:*

1) *We consider the following setup.*

[1]Note that for notational convenience, here we impose a direction on every edge (f_i, f_j), *i.e.*, we consider $i < j$. This inequality is irrelevant for our results, i.e., the results in this paper also hold if we consider $j > i$ for e.

[2]For some special S-NFGs, the associated local functions are different and we use notation different from f.

[3]In this paper, the partition function $Z(\mathsf{N})$ of N is a scalar, *i.e.*, it is not really a function. If N depends on some parameter (say, some temperature parameter), then $Z(\mathsf{N})$ is a function of that parameter.

a) *For each* $t \in \mathbb{Z}_{\geq 0}$*, we consider the following vector of messages and the associated normalization constant:*

$$\boldsymbol{\mu}_{e\to f}^{(t)} \triangleq \left(\mu_{e\to f}^{(t)}(x_e)\right)_{x_e} \in \mathbb{R}_{\geq 0}^{|\mathcal{X}_e|},$$
$$C_{e\to f_i}^{(t)} = \sum_{\boldsymbol{z}_{f_j}} f_j(\boldsymbol{z}_{f_j}) \cdot \prod_{e'\in\partial f_i\setminus\{e\}} \mu_{e'\to f_j}^{(t)}(z_{e',f_j}) \in \mathbb{R}_{\geq 0}, \qquad e=(f_i,f_j)\in\mathcal{E},$$

where $C_{e\to f_j}^{(t)}$ *is defined similarly for each* $e=(f_i,f_j)\in\mathcal{E}$*.*

b) *For* $t=0$*, we randomly generate* $\boldsymbol{\mu}_{e\to f}^{(0)}$ *following the uniform distribution in* $(0,1]^{|\mathcal{X}_e|}$ *for all* $e\in\partial f$*, and* $f\in\mathcal{F}$*.*

2) *We update the messages as follows until some termination criterion is met.*[4]

a) *For every* $t\in\mathbb{Z}_{>0}$ *and* $e=(f_i,f_j)\in\mathcal{E}$ *we first update the normalization constants* $C_{e\to f_i}^{(t-1)}$ *and* $C_{e\to f_j}^{(t-1)}$*. Then we update the messages according to*

$$\mu_{e\to f_i}^{(t)}(x_e) = \frac{1}{C_{e\to f_i}^{(t-1)}} \cdot \sum_{\boldsymbol{z}_{f_j}:\, z_e=x_e} f_j(\boldsymbol{z}_{f_j}) \cdot \prod_{e'\in\partial f_i\setminus\{e\}} \mu_{e'\to f_j}^{(t-1)}(z_{e',f_j}). \tag{1}$$

The collection of messages $\boldsymbol{\mu}_{e\to f_j}^{(t)}$ *is updated similarly.*

3) *For every* $t\in\mathbb{Z}_{>0}$*, the collection of messages* $\boldsymbol{\mu}^{(t)} \triangleq \{\mu_{e\to f}^{(t)}(x_e)\}_{x_e\in\mathcal{X}_e,\, e\in\partial f,\, f\in\mathcal{F}}$ *is called a collection of SPA fixed-point messages if it satisfies*

$$\mu_{e\to f_i}^{(t+1)}(x_e) = \mu_{e\to f_i}^{(t)}(x_e), \qquad x_e\in\mathcal{X}_e,\, e\in\partial f,\, f\in\mathcal{F}.$$

4) *For every* $t\in\mathbb{Z}_{>0}$ *and* $f\in\mathcal{F}$*, the normalization coefficient* $C_f^{(t)}$ *is given by*

$$C_f^{(t)} \triangleq \sum_{\boldsymbol{x}_f} f(\boldsymbol{x}_f) \cdot \prod_{e\in\partial f} \mu_{e\to f}^{(t)}(x_e) \in \mathbb{R}_{\geq 0}.$$

If $C_f^{(t)}>0$*, then the belief obtained by the SPA message* $\boldsymbol{\mu}^{(t)}$ *for function node* f *is given by*

$$\beta_f^{(t)}(\boldsymbol{x}_f) \triangleq \frac{1}{C_f^{(t)}} \cdot f(\boldsymbol{x}_f) \cdot \prod_{e\in\partial f} \mu_{e\to f}^{(t)}(x_e), \quad \boldsymbol{x}_f\in\mathcal{X}_f. \tag{2}$$

For each $e\in\partial f$*, the marginal* $\beta_{f,e}^{(t)}$ *is defined to be*

$$\beta_{f,e}^{(t)}(x_e) \triangleq \sum_{\boldsymbol{z}_f:\, z_e=x_e} \beta_f^{(t)}(\boldsymbol{z}_f), \qquad x_e\in\mathcal{X}_e.$$

5) *For every* $t\in\mathbb{Z}_{>0}$ *and* $e=(f_i,f_j)\in\mathcal{E}$*, the normalization coefficient* $C_e^{(t)}$ *is defined to be*

$$C_e^{(t)} \triangleq \sum_{x_e} \mu_{e\to f_i}^{(t)}(x_e)\cdot\mu_{e\to f_j}^{(t)}(x_e).$$

If $C_e^{(t)}>0$*, the belief obtained by the SPA message* $\boldsymbol{\mu}^{(t)}$ *at edge* e *is defined to be*

$$\beta_e^{(t)}(x_e) \triangleq \frac{1}{C_e^{(t)}} \cdot \mu_{e\to f_i}^{(t)}(x_e)\cdot\mu_{e\to f_j}^{(t)}(x_e), \qquad x_e\in\mathcal{X}_e. \tag{3}$$

At an SPA fixed point, the beliefs $\beta_f^{(t)}$ and $\beta_e^{(t)}$ can be viewed as an approximation of the true marginals $p_{\partial f}$ and $p_{\{e\}}$ induced by the PMF p, respectively, as specified in Definition 3.

Proposition 6. *Given a collection of SPA fixed-point messages* $\boldsymbol{\mu}^{(t)}$*, the beliefs* $\beta_f^{(t)}$ *and* $\beta_e^{(t)}$ *evaluated at the fixed point satisfy*

$$\beta_{f_i,e}^{(t)}(x_e) = \beta_{f_j,e}^{(t)}(x_e) = \beta_e^{(t)}(x_e), \qquad e=(f_i,f_j)\in\mathcal{E}.$$

Proof. It can proven straightforwardly from the definition of the SPA in Definition 5. ■

[4]In general, the termination criterion is a combination of numerical convergence and an upper bound on the number of iterations.

III. The Primal and Dual Formulations of the Bethe Partition Function

In this section, we present the primal and dual formulations of the Bethe partition function for an S-NFG N. The Bethe partition is a function of the minimum of the Bethe free energy function, which can be viewed as an approximation of the partition function of an S-NFG. The primal formulation is mainly based on [9, Section V], and the dual formulation is motivated by the main theoretical results in Heskes' paper [8]. In both these two formulations, some of the stationary points of the objective functions correspond to the SPA fixed points.

1) In the primal formulation, the objective function, i.e., the Bethe free energy function, is neither convex nor concave in general, and the feasible set is convex and formed by some linear constraints.
2) In the dual formulation, the objection function is concave when some of its arguments are fixed, and the feasible set for the arguments has a simple structure. These properties enable us to gain insights for the associated locations of the optimal value and obtain the main results in Section VII.

A. The Primal Formulation

In this subsection, we give the primal formulation of the Bethe partition function. We introduce the local marginal polytope (LMP) for an S-NFG N first.

Definition 7. *Given an S-NFG $\mathsf{N}(\mathcal{F}, \mathcal{E}, \mathcal{X})$, we define*

$$\boldsymbol{\beta} \triangleq (\boldsymbol{\beta}_{\mathcal{E}}, \boldsymbol{\beta}_{\mathcal{F}}), \qquad \boldsymbol{\beta}_{\mathcal{E}} \triangleq \{\boldsymbol{\beta}_e\}_{e\in\mathcal{E}}, \qquad \boldsymbol{\beta}_{\mathcal{F}} \triangleq \{\boldsymbol{\beta}_f\}_{f\in\mathcal{F}},$$
$$\boldsymbol{\beta}_e \triangleq \Big(\beta_e(x_e)\Big)_{x_e\in\mathcal{X}_e} \in \mathbb{R}_{\geq 0}^{|\mathcal{X}_e|}, \qquad \boldsymbol{\beta}_f \triangleq \Big(\beta_f(\boldsymbol{x}_f)\Big)_{\boldsymbol{x}_f\in\mathcal{X}_f} \in \mathbb{R}_{\geq 0}^{|\mathcal{X}_f|}.$$

Note that if $\partial f = \{e_1, e_2\}$ for some edges e_1 and e_2 in $\mathcal{E}$, we can define $\boldsymbol{\beta}_f$ to be a matrix where the rows and columns are indexed by variables x_{e_1} and x_{e_2} respectively. We define

$$\mathcal{B}_f \triangleq \left\{ \boldsymbol{\beta}_f \;\middle|\; \sum_{\boldsymbol{x}_f} \beta_f(\boldsymbol{x}_f) = 1;\, \beta_f(\boldsymbol{x}_f) \in \mathbb{R}_{\geq 0},\, \forall \boldsymbol{x}_f \in \mathcal{X}_f \right\}, \qquad f \in \mathcal{F}, \tag{4}$$

$$\mathcal{B}_e^{\geq} \triangleq \left\{ \boldsymbol{\beta}_e \;\middle|\; \sum_{x_e} \beta_e(x_e) = 1;\, \beta_e(x_e) \in \mathbb{R}_{\geq 0},\, \forall x_e \in \mathcal{X}_e \right\}, \qquad e \in \mathcal{E}, \tag{5}$$

$$\mathcal{B}_e^{>} \triangleq \left\{ \boldsymbol{\beta}_e \;\middle|\; \sum_{x_e} \beta_e(x_e) = 1;\, \beta_e(x_e) \in \mathbb{R}_{> 0},\, \forall x_e \in \mathcal{X}_e \right\}, \qquad e \in \mathcal{E}. \tag{6}$$

Then the LMP is defined to be the set $\mathcal{B}(\mathsf{N})$

$$\mathcal{B}(\mathsf{N}) \triangleq \left(\boldsymbol{\beta} \;\middle|\; \begin{array}{c} \boldsymbol{\beta}_e \in \mathcal{B}_e^{\geq},\, \forall e \in \mathcal{E}, \\ \boldsymbol{\beta}_f \in \mathcal{B}_f,\, \forall f \in \mathcal{F}, \\ \beta_e(z_e) = \beta_{f,e}(x_e),\, \forall f \in \mathcal{F}, e \in \partial f, z_e \in \mathcal{X}_e \text{ (local consistency constraints)} \end{array} \right), \tag{7}$$

where $\boldsymbol{\beta} \in \mathcal{B}(\mathsf{N})$ is called a belief or pseudo-marginal vector, and $\beta_{f,e}$ is the marginal of β_f:

$$\beta_{f,e}(x_e) \triangleq \sum_{\boldsymbol{z}_f :\, z_e = x_e} \beta_f(\boldsymbol{z}_f), \quad x_e \in \mathcal{X}_e,\, e \in \partial f,\, f \in \mathcal{F}. \tag{8}$$

We define an another LMP for $\boldsymbol{\beta}_{\mathcal{F}}$:

$$\mathcal{B}_{\mathcal{F}}(\mathsf{N}) \triangleq \left(\boldsymbol{\beta}_{\mathcal{F}} \;\middle|\; \begin{array}{c} \boldsymbol{\beta}_f \in \mathcal{B}_f,\, \forall f \in \mathcal{F}, \\ \beta_{f_i,e}(x_e) = \beta_{f_j,e}(x_e), \\ x_e \in \mathcal{X}_e,\, e = (f_i, f_j) \in \mathcal{E} \text{ (local consistency constraints)} \end{array} \right). \tag{9}$$

We make some remarks on the above definitions.

- The only difference between $\mathcal{B}_e^{\geq}$ and $\mathcal{B}_e^{>}$ is that in $\mathcal{B}_e^{\geq}$, we consider $\beta_e(x_e) \in \mathbb{R}_{\geq 0}$ for all $x_e \in \mathcal{X}_e$, while in $\mathcal{B}_e^{>}$, we consider $\beta_e(x_e) \in \mathbb{R}_{>0}$ for all $x_e \in \mathcal{X}_e$.
- The condition $\boldsymbol{\beta} = (\boldsymbol{\beta}_{\mathcal{E}}, \boldsymbol{\beta}_{\mathcal{F}}) \in \mathcal{B}(\mathsf{N})$ implies $\boldsymbol{\beta}_{\mathcal{F}} \in \mathcal{B}_{\mathcal{F}}(\mathsf{N})$.
- The sets $\mathcal{B}_f$ and $\mathcal{B}_e^{\geq}$ are sets of vectors representing probability mass functions over $\mathcal{X}_f$ and $\mathcal{X}_e$, respectively.

Definition 8. *[9] The Bethe free energy function w.r.t.* $\mathsf{N}(\mathcal{F}, \mathcal{E}, \mathcal{X})$ *is defined to be the mapping*

$$F_{\mathrm{B,p,N}} : \mathcal{B}(\mathsf{N}) \to \mathbb{R}, \quad \boldsymbol{\beta} \mapsto \sum_f U_{\mathrm{B},f}(\boldsymbol{\beta}_f) - \sum_f H_{\mathrm{B},f}(\boldsymbol{\beta}_f) + \sum_e H_{\mathrm{B},e}(\boldsymbol{\beta}_e), \tag{10}$$

where

$$U_{\mathrm{B},f} : \mathcal{B}_f \to \mathbb{R}, \quad \boldsymbol{\beta}_f \mapsto -\sum_{\boldsymbol{x}_f} \beta_f(\boldsymbol{x}_f) \cdot \log f(\boldsymbol{x}_f),$$

$$H_{\mathrm{B},f} : \mathcal{B}_f \to \mathbb{R}, \quad \boldsymbol{\beta}_f \mapsto -\sum_{\boldsymbol{x}_f} \beta_f(\boldsymbol{x}_f) \cdot \log \beta_f(\boldsymbol{x}_f),$$

$$H_{\mathrm{B},e} : \mathcal{B}_e \to \mathbb{R}, \quad \boldsymbol{\beta}_f \mapsto -\sum_{x_e} \beta_e(x_e) \cdot \log \beta_e(x_e).$$

The letter p *in* $F_{\mathrm{B,p,N}}$ *means that it is related to the primal formulation of the Bethe partition function. We also define the function* $F_{\mathrm{B,p,N}}^{(1)}$ *to be*

$$F_{\mathrm{B,p,N}}^{(1)} : \prod_f \mathcal{B}_f \to \mathbb{R}, \qquad \boldsymbol{\beta}_{\mathcal{F}} \mapsto \sum_f U_{\mathrm{B},f}(\boldsymbol{\beta}_f) - \sum_f H_{\mathrm{B},f}(\boldsymbol{\beta}_f) + \sum_e H_{\mathrm{B},e}\left(\frac{\boldsymbol{\beta}_{f_i,e} + \boldsymbol{\beta}_{f_j,e}}{2}\right). \tag{11}$$

Proposition 9. *If* $\boldsymbol{\beta} \in \mathcal{B}(\mathsf{N})$*, we have* $\boldsymbol{\beta}_{\mathcal{F}} \in \mathcal{B}_{\mathcal{F}}(\mathsf{N})$ *and* $F_{\mathrm{B,p,N}}(\boldsymbol{\beta}) = F_{\mathrm{B,p,N}}^{(1)}(\boldsymbol{\beta}_{\mathcal{F}})$.

Proof. It can be proven straightforwardly following the definitions in Definition 8. ∎

Theorem 10. *The authors of [9] related the fixed points of the SPA for an S-NFG* N *to the stationary points of the Bethe free energy function.*

1) *[9, Theorem 2] If the stationary point of the Bethe free energy function is in the interior of the LMP, i.e.,* $\boldsymbol{\beta} \in \mathcal{B}(\mathsf{N})$ *and* $\boldsymbol{\beta}_f \in \mathbb{R}_{>0}^{|\mathcal{X}_f|}$ *for all* $f \in \mathcal{F}$*, then it corresponds to an SPA fixed points of* N.
2) *[9, Theorem 3] If the S-NFG* N *contains only positive-valued local functions, i.e.,* $f(\boldsymbol{x}_f) \in \mathbb{R}_{>0}$ *for all* $\boldsymbol{x}_f \in \prod_{e \in \partial f} \mathcal{X}_e$ *and* $f \in \mathcal{F}$*, then all local minima of the Bethe free energy function correspond to SPA fixed point of* N.

When the S-NFG N is tree-structured, i.e, cycle-free, the minimum of the Bethe free energy function equals $-\log(Z(\mathsf{N}))$ and the location of the optimal value

$$\boldsymbol{\beta} \in \operatorname{argmin}_{\boldsymbol{\beta} \in \mathcal{B}(\mathsf{N})} F_{\mathrm{B,p,N}}(\boldsymbol{\beta}),$$

satisfies

$$\beta_f(\boldsymbol{x}_f) = p_f(\boldsymbol{x}_f), \qquad \boldsymbol{x}_f \in \mathcal{X}_f,\ f \in \mathcal{F},$$

$$\beta_e(x_e) = p_e(x_e), \qquad x_e \in \mathcal{X}_e,\ e \in \mathcal{E},$$

as proven in [9, Proposition 3] and [17, Theorem 4.2]. For general S-NFG N, the minimum of the Bethe free energy function can be viewed as an approximation of $-\log(Z(\mathsf{N}))$, and the elements $\boldsymbol{\beta}_f$ and $\boldsymbol{\beta}_e$ in the associated the location of the optimal value can be viewed as an approximation of the marginals p_f and p_e induced by the PMF p, respectively.

Definition 11. *The Bethe approximation of the partition function of* N*, i.e., the Bethe partition function, is defined to be*

$$Z_{\mathrm{B,p,N}}^{*} \triangleq \exp\left(-\min_{\boldsymbol{\beta} \in \mathcal{B}(\mathsf{N})} F_{\mathrm{B,p,N}}(\boldsymbol{\beta})\right) = \exp\left(-\min_{\boldsymbol{\beta}_{\mathcal{F}} \in \mathcal{B}_{\mathcal{F}}(\mathsf{N})} F_{\mathrm{B,p,N}}^{(1)}(\boldsymbol{\beta}_{\mathcal{F}})\right), \tag{12}$$

where $Z^{*}_{\mathrm{B,p,N}}$ is the optimal value of the primal formulation of the Bethe partition function. We also define

$$F^{*}_{\mathrm{B,p,N}} \triangleq \min_{\boldsymbol{\beta}\in\mathcal{B}(\mathsf{N})} F_{\mathrm{B,p,N}}(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta}_{\mathcal{F}}\in\mathcal{B}_{\mathcal{F}}(\mathsf{N})} F^{(1)}_{\mathrm{B,p,N}}(\boldsymbol{\beta}_{\mathcal{F}}). \tag{13}$$

In this paper, we mainly focus on analyzing the function $F^{(1)}_{\mathrm{B,p,N}}$ instead of $F_{\mathrm{B,p,N}}$. When $\boldsymbol{\beta}\in\mathcal{B}(\mathsf{N})$, i.e., $\boldsymbol{\beta}_{\mathcal{F}}\in\mathcal{B}_{\mathcal{F}}(\mathsf{N})$, we have $F_{\mathrm{B,p,N}}(\boldsymbol{\beta}) = F_{\mathrm{B,p,N}}(\boldsymbol{\beta}_{\mathcal{F}})$. There are two advantages for analyzing $F^{(1)}_{\mathrm{B,p,N}}$.

- From the definition of $F^{(1)}_{\mathrm{B,p,N}}$ in (11), the associated argument $\boldsymbol{\beta}_{\mathcal{F}}$ is in $\prod_f \mathcal{B}_f$ instead of $\mathcal{B}_{\mathcal{F}}(\mathsf{N})$, while the function $F_{\mathrm{B,p,N}}$ defined in (10) requires that the associated argument $\boldsymbol{\beta}$ is in $\mathcal{B}(\mathsf{N})$. As we will see in Section VII, a key step for obtaining the main results is considering the function $F^{(1)}_{\mathrm{B,p,N}}$ with an argument $\boldsymbol{b}^{(m,n,k)}$ that is allowed to be outside the set $\mathcal{B}_{\mathcal{F}}(\mathsf{N})$. (For details, see (129).)
- We eliminate $\boldsymbol{\beta}_e$ in $F^{(1)}_{\mathrm{B,p,N}}$ by considering $(\boldsymbol{\beta}_{f_i,e}+\boldsymbol{\beta}_{f_j,e})/2$ and the constraint $\boldsymbol{\beta}_{f_i,e}=\boldsymbol{\beta}_{f_j,e}$ instead.

B. Definition of the Dual Formulation

In this section, we present a dual formulation of the Bethe partition function, which provides a different perspective to understand the Bethe partition function. In this paper, we show that the Bethe partition function is equivalent to a maximin problem. The main idea of this transformation was presented in [8, Section 4]. We make a comparison between the results in [8, Section 4] and the results in this paper as follows.

- In [8, Section 4], the saddle-point problem proposed by the author was not well defined, i.e, he used $\max$ and $\min$ in the considered problem, which are indeed $\sup$ and $\inf$, respectively. Also the author did not analyze the locations of the optimal value for the saddle-point problem, i.e, the locations of the optimal value.
- In this paper, we introduce a well-defined maximin problem which is indeed a dual formulation of the Bethe partition function. (For details, see Theorem 57).

The reason why we call the transformed optimization problem the dual formulation of the Bethe partition function is that this transformation process consists of expressing parts of $F_{\mathrm{B,p,N}}$ in terms of their conjugate dual and solving the Lagrangian dual problem of the minimization problem in (13). For details, see the proof of Proposition 16.

Before presenting the dual formulation, we make some definitions.

Definition 12. *We make the following definitions.*

1) *For every edge $e=(f_i,f_j)\in\mathcal{E}(\mathsf{N})$, we define the variables to be*

$$\boldsymbol{\lambda}_e \triangleq \big(\lambda_e(x_e)\big)_{x_e} \in \mathbb{R}^{|\mathcal{X}_e|},$$

$$\boldsymbol{\gamma}_e \triangleq \big(\gamma_e(x_e)\big)_{x_e} \in \mathcal{B}_e^{\geq}, \tag{14}$$

$$\boldsymbol{\lambda}_{e,f_i} \triangleq \boldsymbol{\lambda}_e, \qquad \boldsymbol{\lambda}_{e,f_j} \triangleq -\boldsymbol{\lambda}_e, \qquad 1\leq i<j\leq |\mathcal{F}(\mathsf{N})|, \tag{15}$$

where $\mathcal{B}_e$ is defined in (5).

2) *We define the collections of vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ to be*

$$\boldsymbol{\lambda} \triangleq \{\boldsymbol{\lambda}_e\}_{e\in\mathcal{E}}, \qquad \boldsymbol{\gamma} \triangleq \{\boldsymbol{\gamma}_e\}_{e\in\mathcal{E}}.$$

3) *For every $f\in\mathcal{F}(\mathsf{N})$, we define the collections of vectors $\boldsymbol{\lambda}_{\partial f}$ and $\boldsymbol{\gamma}_{\partial f}$ to be*

$$\boldsymbol{\lambda}_{\partial f} \triangleq (\boldsymbol{\lambda}_{e,f})_{e\in\partial f}, \qquad \boldsymbol{\gamma}_{\partial f} \triangleq (\boldsymbol{\gamma}_e)_{e\in\partial f}.$$

Then we define the function Z_f to be

$$Z_f(\boldsymbol{\gamma}_{\partial f},\boldsymbol{\lambda}_{\partial f}) \triangleq \sum_{\boldsymbol{x}_f} f(\boldsymbol{x}_f)\cdot \prod_{e\in\partial f}\Big(\exp\big(\lambda_{e,f}(x_e)\big)\cdot\sqrt{\gamma_e(x_e)}\Big), \qquad \boldsymbol{x}_f\in\mathcal{X}_f,\ f\in\mathcal{F}. \tag{16}$$

4) *We define the functions* $F_{\mathrm{B,d,N}}$ *and* $Z_{\mathrm{B,d,N}}$ *to be*

$$F_{\mathrm{B,d,N}}(\boldsymbol{\gamma},\boldsymbol{\lambda}) \triangleq -\log\left(\prod_{f} Z_f(\boldsymbol{\gamma}_{\partial f},\boldsymbol{\lambda}_{\partial f})\right) \in \mathbb{R}\cup\{-\infty,\infty\}, \tag{17}$$

$$Z_{\mathrm{B,d,N}}(\boldsymbol{\gamma},\boldsymbol{\lambda}) \triangleq \prod_{f} Z_f(\boldsymbol{\gamma}_{\partial f},\boldsymbol{\lambda}_{\partial f}) \in \mathbb{R}_{\geq 0}\cup\{\infty\}. \tag{18}$$

Then we define $Z^{*}_{\mathrm{B,d,N}}$ *to be the optimal value of a maximin problem:*

$$\begin{aligned} Z^{*}_{\mathrm{B,d,N}} \triangleq \sup_{\boldsymbol{\gamma}} \inf_{\boldsymbol{\lambda}} \, & Z_{\mathrm{B,d,N}}(\boldsymbol{\gamma},\boldsymbol{\lambda}) \\ \text{s.t.} \quad & \lambda_e(x_e)\in\mathbb{R},\, x_e\in\mathcal{X}_e,\, \boldsymbol{\gamma}_e\in\mathcal{B}_e^{>},\, e\in\mathcal{E}, \end{aligned} \tag{19}$$

where $\mathcal{B}_e^{>}$ *is defined in* (6)*. The reason why we use* $\sup_{\boldsymbol{\gamma}}$ *instead of* $\max_{\boldsymbol{\gamma}}$ *is that the function* $\inf_{\boldsymbol{\lambda}} Z_{\mathrm{B,d,N}}(\boldsymbol{\gamma},\boldsymbol{\lambda})$ *maybe discontinuous w.r.t.* $\boldsymbol{\gamma}\in\prod_e \mathcal{B}_e^{>}$*. Note that the letter* d *in* $Z^{*}_{\mathrm{B,d,N}}$ *means that it is the optimal value of the dual formulation of the Bethe partition function, which will be proven in Proposition 16. We also define* $F^{*}_{\mathrm{B,d,N}}$ *to be the optimal value of a minimax problem:*

$$\begin{aligned} F^{*}_{\mathrm{B,d,N}} \triangleq \inf_{\boldsymbol{\gamma}} \sup_{\boldsymbol{\lambda}} \, & F_{\mathrm{B,d,N}}(\boldsymbol{\gamma},\boldsymbol{\lambda}) \\ \text{s.t.} \quad & \lambda_e(x_e)\in\mathbb{R},\, x_e\in\mathcal{X}_e,\, \boldsymbol{\gamma}_e\in\mathcal{B}_e^{>},\, e\in\mathcal{E}. \end{aligned} \tag{20}$$

5) *We consider another optimization problem that is closely related to* $Z^{*}_{\mathrm{B,d,N}}$*:*

$$\begin{aligned} Z^{\mathrm{alt},*}_{\mathrm{B,d,N}} \triangleq \sup_{\boldsymbol{\gamma}} \inf_{\boldsymbol{\lambda}} \, & Z_{\mathrm{B,d,N}}(\boldsymbol{\gamma},\boldsymbol{\lambda}) \\ \text{s.t.} \quad & \lambda_e(x_e)\in\mathbb{R},\, x_e\in\mathcal{X}_e,\, \boldsymbol{\gamma}_e\in\mathcal{B}_e^{\geq},\, e\in\mathcal{E}, \end{aligned} \tag{21}$$

where $\mathcal{B}_e^{\geq}$ *is defined in* (5)*. Similarly, we define* $F^{\mathrm{alt},*}_{\mathrm{B,d,N}}$ *to be*

$$\begin{aligned} F^{\mathrm{alt},*}_{\mathrm{B,d,N}} \triangleq \inf_{\boldsymbol{\gamma}} \sup_{\boldsymbol{\lambda}} \, & F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma},\boldsymbol{\lambda}) \\ \text{s.t.} \quad & \lambda_e(x_e)\in\mathbb{R},\, x_e\in\mathcal{X}_e,\, \boldsymbol{\gamma}_e\in\mathcal{B}_e^{\geq},\, e\in\mathcal{E}, \end{aligned} \tag{22}$$

where

$$F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma},\boldsymbol{\lambda}) \triangleq F_{\mathrm{B,d,N}}(\boldsymbol{\gamma},\boldsymbol{\lambda}) = -\log\left(\prod_{f} Z_f(\boldsymbol{\gamma}_{\partial f},\boldsymbol{\lambda}_{\partial f})\right). \tag{23}$$

The only difference between $Z^{\mathrm{alt},*}_{\mathrm{B,d,N}}$ *and* $Z^{*}_{\mathrm{B,d,N}}$ *is that in* $Z^{*}_{\mathrm{B,d,N}}$*, we consider the vector* $\boldsymbol{\gamma}_e$ *in* $\mathcal{B}_e^{>}$ *for all* e *in* $\mathcal{E}$*, while in* $Z^{\mathrm{alt},*}_{\mathrm{B,d,N}}$*, we consider the vector* $\boldsymbol{\gamma}_e$ *in* $\mathcal{B}_e^{\geq}$ *for all* e *in* $\mathcal{E}$*. The difference between* $F^{\mathrm{alt},*}_{\mathrm{B,d,N}}$ *and* $F^{*}_{\mathrm{B,d,N}}$ *is the same as the difference between* $Z^{\mathrm{alt},*}_{\mathrm{B,d,N}}$ *and* $Z^{*}_{\mathrm{B,d,N}}$*.*

6) *For each* $f\in\mathcal{F}$*, we define the belief based on the variables* $\boldsymbol{\lambda}_{\partial f}$ *and* $\boldsymbol{\gamma}_{\partial f}$ *to be*

$$b_f(\boldsymbol{x}_f) \triangleq \frac{f(\boldsymbol{x}_f)\cdot\prod_{e\in\partial f}\Big(\exp\big(\lambda_{e,f}(x_e)\big)\cdot\sqrt{\gamma_e(x_e)}\Big)}{Z_f\big(\boldsymbol{\gamma}_{\partial f},\boldsymbol{\lambda}_{\partial f}\big)} \in \mathcal{B}_f, \qquad Z_f\big(\boldsymbol{\gamma}_{\partial f},\boldsymbol{\lambda}_{\partial f}\big) > 0. \tag{24}$$

Let $\mathcal{I}_f$ *be a subset of* ∂f *and let* $\mathcal{I}_f^{\mathrm{c}} \triangleq \partial f\setminus\mathcal{I}_f$ *be its complement. The marginal* $b_{f,\mathcal{I}_f}(\boldsymbol{x}_{\mathcal{I}_f})$ *is defined to be*

$$b_{f,\mathcal{I}_f}(\boldsymbol{x}_{\mathcal{I}_f}) \triangleq \sum_{\boldsymbol{x}_{\mathcal{I}_f^{\mathrm{c}}}\in\mathcal{X}_{\mathcal{I}_f^{\mathrm{c}}}} b_f(\boldsymbol{x}_f),$$

where

$$\mathcal{X}_{\mathcal{I}_f} \triangleq \prod_{e\in\mathcal{I}_f}\mathcal{X}_e, \qquad \mathcal{X}_{\mathcal{I}_f^{\mathrm{c}}} \triangleq \prod_{e\in\mathcal{I}_f^{\mathrm{c}}}\mathcal{X}_e,$$

$$\boldsymbol{x}_{\mathcal{I}_f} \triangleq (x_e)_{e\in\mathcal{I}_f}\in\mathcal{X}_{\mathcal{I}_f}, \qquad \boldsymbol{x}_{\mathcal{I}_f^{\mathrm{c}}} \triangleq (x_e)_{e\in\mathcal{I}_f^{\mathrm{c}}}\in\mathcal{X}_{\mathcal{I}_f^{\mathrm{c}}}.$$

For example, the function $b_{f,\{e\}}$ is given by

$$b_{f,\{e\}}(x_e) = \sum_{\boldsymbol{z}_f:\, z_e = x_e} b_f(\boldsymbol{z}_f), \qquad e \in \partial f,\, f \in \mathcal{F}. \tag{25}$$

In the remaining part of this paper, we use the shorthand $b_{f,e}$ for $b_{f,\{e\}}$ for simplicity.

Similarly, we can relate some of the stationary points of the objective function in the optimization problem (22) to the SPA fixed points.

Proposition 13. *The following two statements hold.*

1) *The optimization problem* (22) *is equivalent to the following optimization problem*

$$\begin{aligned} F_{\mathrm{B,d,N}}^{\mathrm{alt},*} = \inf_{\boldsymbol{\gamma}} \sup_{\boldsymbol{\lambda}} & \left\{ F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) + \sum_e \log\left(\sum_{x_e} \gamma_e(x_e) \right) \right\} \\ \text{s.t.} \quad & \lambda_e(x_e) \in \mathbb{R},\, \gamma_e(x_e) \in \mathbb{R}_{\geq 0},\, x_e \in \mathcal{X}_e,\, \sum_{x_e} \gamma_e(x_e) \in \mathbb{R}_{>0},\, e \in \mathcal{E}. \end{aligned} \tag{26}$$

   Note that compared with the optimization problem in (22)*, the constraints $\boldsymbol{\gamma}_e \in \mathcal{B}_e^{\geq}$ for all $e \in \mathcal{E}$, were changed into the constraints $\gamma_e(x_e) \in \mathbb{R}_{\geq 0}$.*

2) *If there are variables $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ such that*
   a) *$\boldsymbol{\gamma}_e \in \mathcal{B}_e^{>}$ for all $e \in \mathcal{E}$;*
   b) *$\boldsymbol{\lambda} \in \mathbb{R}^{|\mathcal{X}|}$;*
   c) *$\frac{\partial}{\partial \gamma_e(x_e)} F_{\mathrm{B,d,N}}^{\mathrm{alt}} + 1 = 0$ and $\frac{\partial}{\partial \lambda_e(x_e)} F_{\mathrm{B,d,N}}^{\mathrm{alt}} = 0$ for all $x_e \in \mathcal{X}_e$ and $e \in \mathcal{E}$, i.e., the collection of vectors $(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ is at the stationary point of the objective function in the optimization problem* (26)*,*

   then $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ correspond to an SPA fixed point.

Proof. See Appendix A. ∎

Let us discuss some properties of the scalars $F_{\mathrm{B,d,N}}^{*}$, $F_{\mathrm{B,d,N}}^{\mathrm{alt},*}$, and the function $F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ as specified in Definition 12.

Theorem 14. *[8], [16] The functions $F_{\mathrm{B,d,N}}$ and $F_{\mathrm{B,d,N}}^{\mathrm{alt}}$ defined in* (17) *and* (23)*, respectively, have the following properties.*

1) *[8, Section 4] For fixed $\boldsymbol{\gamma}_e \in \mathcal{B}_e^{\geq}$ for all $e \in \mathcal{E}$, both the functions $F_{\mathrm{B,d,N}}$ and $F_{\mathrm{B,d,N}}^{\mathrm{alt}}$ are concave w.r.t. $\boldsymbol{\lambda}$.*
2) *[16, Section 6.2] For fixed $\boldsymbol{\lambda}$ and $\{\boldsymbol{\gamma}_{e'}\}_{e' \in \mathcal{E} \setminus \{e\}}$, both the functions $F_{\mathrm{B,d,N}}$ and $F_{\mathrm{B,d,N}}^{\mathrm{alt}}$ are convex w.r.t. $\boldsymbol{\gamma}_e$.*

Proof. See Appendix B. ∎

Proposition 15. *It holds that*

$$F_{\mathrm{B,d,N}}^{\mathrm{alt},*} \leq F_{\mathrm{B,d,N}}^{*}, \qquad Z_{\mathrm{B,d,N}}^{\mathrm{alt},*} \geq Z_{\mathrm{B,d,N}}^{*}.$$

Proof. As mentioned in Item 5 in Definition 12, the only difference between $F_{\mathrm{B,d,N}}^{*}$ and $F_{\mathrm{B,d,N}}^{\mathrm{alt},*}$ is that

- in $F_{\mathrm{B,d,N}}^{*}$, we consider $\boldsymbol{\gamma}_e$ in $\mathcal{B}_e^{>}$ for all $e \in \mathcal{E}$;
- in $F_{\mathrm{B,d,N}}^{\mathrm{alt},*}$, we consider $\boldsymbol{\gamma}_e$ in $\mathcal{B}_e^{\geq}$ for all $e \in \mathcal{E}$.

Then the proposition follows immediately from the fact that $\mathcal{B}_e^{>}$ is a subset of $\mathcal{B}_e^{\geq}$ for all $e \in \mathcal{E}$. Finally, comparing the definitions of $Z_{\mathrm{B,d,N}}^{*}$ and $Z_{\mathrm{B,d,N}}^{\mathrm{alt},*}$ in (19) and (21), respectively, with the definitions of $F_{\mathrm{B,d,N}}^{*}$ and $F_{\mathrm{B,d,N}}^{\mathrm{alt},*}$ in (20) and (22), respectively, proving $Z_{\mathrm{B,d,N}}^{\mathrm{alt},*} \geq Z_{\mathrm{B,d,N}}^{*}$ is equivalent to proving $F_{\mathrm{B,d,N}}^{\mathrm{alt},*} \leq F_{\mathrm{B,d,N}}^{*}$.

∎

Now we show that the optimization problem in (19) with optimal value $Z^{*}_{\mathrm{B,d,N}}$ is a dual formulation of the Bethe partition $Z^{*}_{\mathrm{B,p,N}}$ function defined in (12). Another dual formulation is the optimization problem in (21) with optimal value $Z^{\mathrm{alt},*}_{\mathrm{B,d,N}}$, which will be proven in Theorem 57.

Proposition 16. *It holds that*

$$Z^{*}_{\mathrm{B,p,N}} = Z^{*}_{\mathrm{B,d,N}}, \tag{27}$$

or, equivalently,

$$F^{*}_{\mathrm{B,p,N}} = F^{*}_{\mathrm{B,d,N}}.$$

It also holds that

$$\Big(\inf_{\boldsymbol{\lambda}} Z_{\mathrm{B,d,N}}(\boldsymbol{\gamma},\boldsymbol{\lambda})\Big) \in \mathbb{R}_{>0}, \qquad \forall \gamma_e(x_e) \in \mathbb{R}_{>0},\, x_e \in \mathcal{X}_e, \quad \sum_{x_e} \gamma_e(x_e) = 1,\, e \in \mathcal{E}. \tag{28}$$

Proof. We start by giving an outline of the proof for (27).

- Recall that the optimal value $Z^{*}_{\mathrm{B,p,N}}$ of the primal formulation, as defined in (12), is related to the minimum of $F_{\mathrm{B,p,N}}$, which is defined in (10) and consists of the finite sum of the entropy functions in $\{H_{\mathrm{B},e}\}_e$. The function $H_{\mathrm{B},e}$ is a convex function whose convex conjugate is a "log-sum-exp" function. We transform the function $H_{\mathrm{B},e}$ into a "log-sum-exp" function for each $e \in \mathcal{E}$ in $F_{\mathrm{B,p,N}}$.
- After that, we minimize the transformed function w.r.t. $\boldsymbol{\beta}_{\mathcal{F}}$ only. Solving this minimization problem is equivalent to solving the convex conjugate of another entropy function, and the resulting function is again a sum of "log-sum-exp" functions. After that, we obtain the dual formulation of the Bethe partition function.

For details, see the proof in Appendix C. ■

In the following, for simplicity, when we talk about the primal formulation and the dual formulation, we mean the formulation of the Bethe partition function and the dual formulation of the Bethe partition function, respectively.

We conclude this section by pointing out that the dualization that is used here is different from the approach in [9], [14], [15]. The differences are sketched in Fig. 2:

- Fig. 2(a) shows parts of an S-NFG of interest.
- Fig. 2(b) shows parts of an NFG whose global function is equal to the Bethe free energy function of the S-NFG in Fig. 2(a). Note that here the global function is the finite *sum* of the local functions, not the product of the local functions.
- Dualizing the S-NFG in Fig. 2(b) according to [9], [14], [15] yields an NFG as in Fig. 2(d).
- The approach by Heskes [8] can be seen as first modifying Fig. 2(b) to obtain Fig. 2(c). Namely, the equal-contraint function node in the middle of Fig. 2(b) is replaced by the, functionally equivalent, dashed box in Fig. 2(c). The NFG in Fig. 2(c) is then dualized and yields the NFG in Fig. 2(e).

C. The Locations of the optimal value for the Dual Formulation

In this section, we study the location of optimal value of the optimization problem in (22). As we will see in Theorem 57, the optimization problem in (22) is indeed the dual formulation of the Bethe partition function. There are two main goals in this paper.

1) We want to prove that $F^{\mathrm{alt},*}_{\mathrm{B,d,N}} = F^{*}_{\mathrm{B,d,N}} = F^{*}_{\mathrm{B,p,N}}$. By Propositions 15 and 16, we have $F^{\mathrm{alt},*}_{\mathrm{B,d,N}} \leq F^{*}_{\mathrm{B,d,N}} = F^{*}_{\mathrm{B,p,N}}$. Therefore, it remains to prove $F^{\mathrm{alt},*}_{\mathrm{B,d,N}} \geq F^{*}_{\mathrm{B,d,N}}$, which will be done in Theorem 57.
2) We want to show that there exists a sequence of messages defined based on the locations of the optimal value of the optimization problem in (22) such that the message sequence converges to an SPA fixed point.

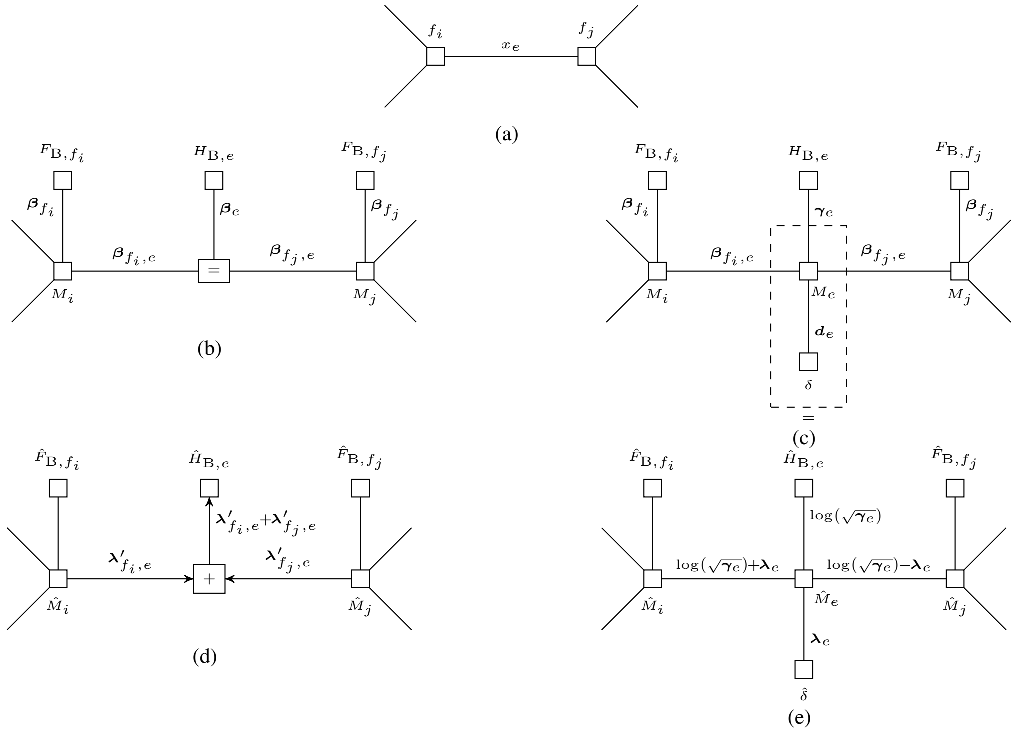


Fig. 2: Different dualizations of the Bethe free energy function.

To achieve these two goals, we first define the sequences based on the locations of the optimal value of the optimization problem in (22).

Definition 17. *We make the following definitions.*

1) *Given $\boldsymbol{\gamma}_e$ in $\mathcal{B}_e^{\geq}$ for each $e \in \mathcal{E}$, we define*

$$\hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}) \triangleq \sup_{\boldsymbol{\lambda}} F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}), \tag{29}$$

where $\mathcal{B}_e^{\geq}$ is defined in (5).

2) *We define $\{\boldsymbol{\gamma}^{(m)}\}_{m \in \mathbb{Z}_{>0}}$ to be a sequence satisfying*

   a) *for each $e \in \mathcal{E}$ and $m \in \mathbb{Z}_{>0}$, the element $\boldsymbol{\gamma}_e^{(m)}$ is in $\mathcal{B}_e^{\geq}$, where $\mathcal{B}_e^{\geq}$ is defined in (5).*

   b) *the associated sequence $\{\hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)})\}_m$ converges to $F_{\mathrm{B,d,N}}^{\mathrm{alt},*}$:*

$$\lim_{m \to \infty} \hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}) = F_{\mathrm{B,d,N}}^{\mathrm{alt},*}. \tag{30}$$

3) *Given $\boldsymbol{\gamma}^{(m)}$, we define the sequence $\left\{\boldsymbol{\lambda}^{(n(m))}\big(\boldsymbol{\gamma}^{(m)}\big)\right\}_{n(m) \in \mathbb{Z}_{>0}}$ to be a sequence such that it converges to one of the locations of the optimal value for the optimization problem in (29). There are two cases to be considered.*

   a) *If there exists $\boldsymbol{\lambda}^*\big(\boldsymbol{\gamma}^{(m)}\big) \in \mathbb{R}^{|\mathcal{X}|}$ such that*

$$F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\mathsf{N}, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\lambda}^*\big(\boldsymbol{\gamma}^{(m)}\big)\Big) = \sup_{\boldsymbol{\lambda}} F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}, \boldsymbol{\lambda}),$$

   then we define

$$\boldsymbol{\lambda}^{(n(m))}\big(\boldsymbol{\gamma}^{(m)}\big) = \boldsymbol{\lambda}^*\big(\boldsymbol{\gamma}^{(m)}\big), \qquad n(m) \in \mathbb{Z}_{>0}.$$

   b) *If not, we define $\left\{\boldsymbol{\lambda}^{(n(m))}\big(\boldsymbol{\gamma}^{(m)}\big)\right\}_{n(m) \in \mathbb{Z}_{>0}}$ to be a sequence satisfying*

$$F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)}, \boldsymbol{\lambda}^{(n(m))}\big(\boldsymbol{\gamma}^{(m)}\big)\Big) \geq F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}, \mathbf{0}), \tag{31}$$

$$\lim_{n(m)\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}\Big(\boldsymbol{\gamma}^{(m)}, \boldsymbol{\lambda}^{(n(m))}\big(\boldsymbol{\gamma}^{(m)}\big)\Big) = \hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}). \tag{32}$$

For simplicity, if there is no ambiguity, in the following we use $\boldsymbol{\lambda}^{(n)}$ and n instead of $\boldsymbol{\lambda}^{(n(m))}\big(\boldsymbol{\gamma}^{(m)}\big)$ and $n(m)$, respectively.

4) *For each n and m, we consider the following optimization problem*

$$\alpha^* \in \operatorname{argmax}_{\alpha\in\mathbb{R}\cup\{-\infty,+\infty\}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha\cdot\boldsymbol{\lambda}^{(n)}). \tag{33}$$

To define a sequence that converges to α^*, there are two cases to be considered.

a) *If there exists an α^* in $\mathbb{R}$, we define the sequence $\{\alpha^{(k(m,n))}\}_{k(m,n)}$ to be*

$$\alpha^{(k(m,n))} \triangleq \alpha^*, \qquad k(m,n)\in\mathbb{Z}_{>0}. \tag{34}$$

For example, if case 3a happens, then we have $\alpha^* = 1$ and $\alpha^{(k(m,n))} = 1$ for all $k(m,n)\in\mathbb{Z}_{>0}$.

b) *Otherwise, we define the sequence $\{\alpha^{(k(m,n))}\}_{k(m,n)}$ to be*

$$\alpha^{(k(m,n))} \triangleq \begin{cases} 2^{k(m,n)} & \alpha^* = +\infty \\ -2^{k(m,n)} & \alpha^* = -\infty \end{cases}, \qquad k(m,n)\in\mathbb{Z}_{>0},$$

where the integer $k(m,n)$ satisfies

$$F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k(m,n))}\cdot\boldsymbol{\lambda}^{(n)}) \geq F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \mathbf{0}), \qquad k(m,n)\in\mathbb{Z}_{>0}. \tag{35}$$

For simplicity, if there is no ambiguity, in the following we use $\alpha^{(k)}$ and k instead of $\alpha^{(k(m,n))}$ and $k(m,n)$, respectively.

5) *For each $f\in\mathcal{F}$, we define the collection of variables $\boldsymbol{x}_{\partial f,f}$ to be*

$$\boldsymbol{x}_{\partial f,f} \triangleq (x_{e,f})_{e\in\partial f} \in \mathcal{X}_f. \tag{36}$$

We also define a sequence of beliefs on function node f based on the sequences $\{\boldsymbol{\gamma}^{(m)}\}_{m\in\mathbb{Z}_{>0}}$, $\{\boldsymbol{\lambda}^{(n)}\}_{n\in\mathbb{Z}_{>0}}$, and $\{\alpha^{(k)}\}_{k\in\mathbb{Z}_{>0}}$. The belief sequence on function node f is given by

$$b_f^{(m,n,k)}(\boldsymbol{x}_f) \triangleq \frac{f(\boldsymbol{x}_f)\cdot\prod_{e\in\partial f}\Big(\exp\Big(\alpha^{(k)}\cdot\lambda_{e,f}^{(n)}(x_e)\Big)\cdot\sqrt{\gamma_e^{(m)}(x_e)}\Big)}{Z_f\big(\boldsymbol{\gamma}_{\partial f}^{(m)}, \alpha^{(k)}\cdot\boldsymbol{\lambda}_{\partial f}^{(n)}\big)}, \qquad \boldsymbol{x}_f\in\mathcal{X}_f,\ Z_f\big(\boldsymbol{\gamma}_{\partial f}^{(m)}, \alpha^{(k)}\cdot\boldsymbol{\lambda}_{\partial f}^{(n)}\big) > 0, \tag{37}$$

where

$$\alpha^{(k)}\cdot\boldsymbol{\lambda}_{\partial f}^{(n)} = \Big(\alpha^{(k)}\cdot\lambda_{e,f}^{(n)}(x_e)\Big)_{x_e\in\mathcal{X}_e,\, e\in\partial f,\, f\in\mathcal{F}} \in \mathbb{R}^{|\mathcal{X}|},$$

$$Z_f\big(\boldsymbol{\gamma}_{\partial f}^{(m)}, \alpha^{(k)}\cdot\boldsymbol{\lambda}_{\partial f}^{(n)}\big) = \sum_{\boldsymbol{x}_f} f(\boldsymbol{x}_f)\cdot\prod_{e\in\partial f}\Big(\exp\Big(\alpha^{(k)}\cdot\lambda_{e,f}^{(n)}(x_e)\Big)\cdot\sqrt{\gamma_e^{(m)}(x_e)}\Big).$$

Then we define the collections of variables $\boldsymbol{b}_f^{(m,n,k)}$ and $\boldsymbol{b}^{(m,n,k)}$ to be

$$\boldsymbol{b}_f^{(m,n,k)} \triangleq \big(b_f^{(m,n,k)}(\boldsymbol{x}_f)\big)_{\boldsymbol{x}_f\in\mathcal{X}_f},$$
$$\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)} \triangleq \big(\boldsymbol{b}_f^{(m,n,k)}\big)_{f\in\mathcal{F}}. \tag{38}$$

6) *Let $\mathcal{I}_f$ be a subset of ∂f and let $\mathcal{I}_f^{\mathrm{c}} = \partial f\setminus\mathcal{I}_f$ be its complement. We define the marginal*

$$b_{f,\mathcal{I}_f}^{(m,n,k)}(\boldsymbol{x}_{\mathcal{I}_f}) \triangleq \sum_{\boldsymbol{x}_{\mathcal{I}_f^{\mathrm{c}}}\in\mathcal{X}_{\mathcal{I}_f^{\mathrm{c}}}} b_f^{(m,n,k)}(\boldsymbol{x}_f), \qquad \boldsymbol{x}_{\mathcal{I}_f}\in\mathcal{X}_{\mathcal{I}_f}. \tag{39}$$

7) *For simplicity, if there is no ambiguity, we used the shorthands $\{\cdot\}_m$, $\{\cdot\}_n$, $\{\cdot\}_k$, and $\{\cdot\}_{n,k}$ for $\{\cdot\}_{m\in\mathbb{Z}_{>0}}$, $\{\cdot\}_{n(m)\in\mathbb{Z}_{>0}}$, $\{\cdot\}_{k(n(m),m)\in\mathbb{Z}_{>0}}$, and $\{\cdot\}_{n(m),k(n(m),m)\in\mathbb{Z}_{>0}}$, respectively. We define the following operators*

$$\lim_{m,n,k\to\infty} \triangleq \lim_{m\to\infty}\lim_{n(m)\to\infty}\lim_{k(m,n)\to\infty},$$

$$\lim_{m,n\to\infty} \triangleq \lim_{m\to\infty} \lim_{n(m)\to\infty},$$

$$\lim_{n,k\to\infty} \triangleq \lim_{n(m)\to\infty} \lim_{k(m,n)\to\infty}.$$

For fixed $\boldsymbol{\gamma}$, as proven in Theorem 14, the function $F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma},\boldsymbol{\lambda})$ is concave w.r.t. $\boldsymbol{\lambda}$. For each $\boldsymbol{\gamma}$, we can solve $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma})$ as defined in (29) by finding the stationary point of $F^{\mathrm{alt}}_{\mathrm{B,d,N_2}}(\boldsymbol{\gamma},\boldsymbol{\lambda})$, which satisfies

$$\frac{\partial}{\partial\lambda_e(x_e)} F^{\mathrm{alt}}_{\mathrm{B,d,N}} = -b_{f_i,e}(x_e) + b_{f_j,e}(x_e) = 0, \qquad x_e \in \mathcal{X}_e,\, e=(f_i,f_j)\in\mathcal{E}, \tag{40}$$

where $b_{f_i,e}$ and $b_{f_j,e}$ are given in (25). If

$$\Big(b_{f_i,e}(x_e), b_{f_j,e}(x_e)\Big) \neq (0,0), \qquad x_e \in \mathcal{X}_e, e\in\mathcal{E},$$

then the equation in (40) implies

$$\exp\bigl(\lambda_e(x_e)\bigr) = \sqrt{\frac{Z_{f_i}(\boldsymbol{\gamma}_{\partial f},\boldsymbol{\lambda}_{\partial f})\cdot \sum\limits_{\boldsymbol{z}_{f_j}:\, z_e=x_e} f_j(\boldsymbol{z}_{f_j})\cdot \prod_{e'\in\partial f_j\setminus\{e\}} \exp\bigl(\lambda_{e',f_j}(z_{e'})\bigr)\cdot\sqrt{\gamma_{e'}(z_{e'})}}{Z_{f_j}(\boldsymbol{\gamma}_{\partial f},\boldsymbol{\lambda}_{\partial f})\cdot \sum\limits_{\boldsymbol{z}_{f_i}:\, z_e=x_e} f_i(\boldsymbol{z}_{f_i})\cdot \prod_{e''\in\partial f_i\setminus\{e\}} \exp\bigl(\lambda_{e'',f_i}(z_{e''})\bigr)\cdot\sqrt{\gamma_{e''}(z_{e''})}}} \in \mathbb{R}_{\geq 0}\cup\{\infty\}, \tag{41}$$

$x_e \in \mathcal{X}_e$, $e\in\mathcal{E}$,

where Z_f is defined in (16) for all $f\in\mathcal{F}$. Note that $\lambda_e(x_e)$ does not appear on the righthand side of (41).

D. Algorithms for Solving the Dual Formulation

In this section, we present a double-loop algorithm in [8, Algorithm 2] for solving the minimax optimization problem in (22) by rewriting the details based on our definitions in the previous sections. The main idea is unchanged.

Definition 18. *[8, Algorithm 2] The details of the algorithm are given in Algorithm 1. The algorithm consists of two loops.*

1) *In the inner loop, we fix $\boldsymbol{\gamma}$ and find the stationary point of $F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma},\boldsymbol{\lambda})$ w.r.t. $\boldsymbol{\lambda}$.*
2) *In the outer loop, we fix $\boldsymbol{\lambda}$ and find the stationary point of $F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma},\boldsymbol{\lambda})$ w.r.t. $\boldsymbol{\gamma}$.*

Then we consider the following setup.

1) *Let $t_1, t_2 \in \mathbb{Z}_{\geq 0}$ be the iteration indices.*
2) *For each $e=(f_i,f_j)\in\mathcal{E}$, we define*

$$\boldsymbol{\lambda}^{(t_1)}_{\mathrm{dl},e} \triangleq \Bigl(\lambda^{(t_1)}_{\mathrm{dl},e}(x_e)\Bigr)_{x_e}, \qquad \boldsymbol{\gamma}^{(t_2)}_{\mathrm{dl},e} \triangleq \Bigl(\gamma^{(t_2)}_{\mathrm{dl},e}(x_e)\Bigr)_{x_e}, \tag{42}$$

$$\boldsymbol{\lambda}^{(t_1)}_{\mathrm{dl},e,f_i} \triangleq \boldsymbol{\lambda}^{(t_1)}_{\mathrm{dl},e}, \qquad \boldsymbol{\lambda}^{(t_1)}_{\mathrm{dl},e,f_j} \triangleq -\boldsymbol{\lambda}^{(t_1)}_{\mathrm{dl},e}. \tag{43}$$

We also define the message sequence based on the above-defined sequences to be the sequence $\{\boldsymbol{\mu}^{(t_1,t_2)}_{e\to f}\}_{t_1,t_2}$ satisfying

$$\mu^{(t_1,t_2)}_{e\to f}(x_e) \propto \exp\Bigl(\lambda^{(t_1)}_{\mathrm{dl},e,f}(x_e)\Bigr)\cdot\sqrt{\gamma^{(t_2)}_{\mathrm{dl},e}(x_e)}, \qquad x_e\in\mathcal{X}_e, \tag{44}$$

$$\sum_{x_e} \mu^{(t_1,t_2)}_{e\to f}(x_e) = 1, \qquad e\in\mathcal{E}, f\in\mathcal{F}. \tag{45}$$

3) *Then we define the following collections of variables:*

$$\boldsymbol{\lambda}^{(t_1)}_{\mathrm{dl}} \triangleq \Bigl(\boldsymbol{\lambda}^{(t_1)}_{\mathrm{dl},e}\Bigr)_{e\in\mathcal{E}}, \qquad \boldsymbol{\gamma}^{(t_2)}_{\mathrm{dl}} \triangleq \Bigl(\boldsymbol{\gamma}^{(t_2)}_{\mathrm{dl},e}\Bigr)_{e\in\mathcal{E}}, \qquad \boldsymbol{\mu}^{(t_1,t_2)} \triangleq \Bigl(\mu^{(t_1,t_2)}_{e\to f}(x_e)\Bigr)_{x_e\in\mathcal{X}_e,\, e\in\partial f,\, f\in\mathcal{F}}.$$

4) *Given $\boldsymbol{\lambda}^{(t_1)}_{\mathrm{dl}}$ and $\boldsymbol{\gamma}^{(t_1)}_{\mathrm{dl}}$, for each $f\in\mathcal{F}$, we define the normalization constant $Z^{(t_1,t_2)}_{\mathrm{dl},f}$ to be*

$$Z^{(t_1,t_2)}_{\mathrm{dl},f} \triangleq \sum_{\boldsymbol{x}_f} f(\boldsymbol{x}_f)\cdot\prod_{e\in\partial f}\mu^{(t_1,t_2)}_{e\to f}(x_e).$$

If $Z_{\mathrm{dl},f}^{(t_1,t_2)}$ is positive-valued, the associated belief on function node f is

$$b_{\mathrm{dl},f}^{(t_1,t_2)}(\boldsymbol{x}_f) \triangleq \frac{1}{Z_{\mathrm{dl},f}^{(t_1,t_2)}} \cdot f(\boldsymbol{x}_f) \cdot \prod_{e\in\partial f} \mu_{e\to f}^{(t_1,t_2)}(x_e), \qquad \boldsymbol{x}_f \in \mathcal{X}_f,\ f\in\mathcal{F}.$$

The associated marginal is defined to be

$$b_{\mathrm{dl},f,e}^{(t_1,t_2)}(x_e) \triangleq \sum_{\boldsymbol{z}_f:\, z_e=x_e} b_{\mathrm{dl},f,e}^{(t_1,t_2)}(\boldsymbol{z}_f), \qquad x_e\in\mathcal{X}_e,\ e\in\partial f.$$

Note the letters "dl" in the above defined variables stands for double loops. The main idea of the algorithm is given as follows. For details, see Algorithm 1.

5) *We randomly generate $\boldsymbol{\gamma}_{\mathrm{dl},e}^{(0)}$ following the uniform distribution in $(0,1]^{|\mathcal{X}_e|}$ for all $e\in\partial f$, and $f\in\mathcal{F}$.*
6) *We randomly generate $\boldsymbol{\lambda}_{\mathrm{dl}}^{(0)}$ following the uniform distribution in $[-1,1]^{|\mathcal{X}_e|}$ for all $e\in\partial f$, and $f\in\mathcal{F}$.*
7) *Fixing the index of the outer loop t_2 and the vector $\boldsymbol{\gamma}_{\mathrm{dl}}^{(t_2)}$, we update $\boldsymbol{\lambda}_{\mathrm{dl}}^{(t_1)}$ in the inner loop until some termination criterion is met.*[5]
   a) *If both $Z_{\mathrm{dl},f_j}^{(t_1-1,t_2-1)}$ and $Z_{\mathrm{dl},f_i}^{(t_1-1,t_2-1)}$ are positive-valued, by the derivations in* (40)–(41)*, to find the stationary point of $F_{\mathrm{B,d,N_1}}^{\mathrm{alt},*}$ w.r.t. $\boldsymbol{\lambda}$, we update variable $\lambda_{\mathrm{dl},e}^{(t_1)}$ such that*

$$\lambda_{\mathrm{dl},e}^{(t_1)}(x_e) \propto \begin{cases} \lambda_{\mathrm{dl},e}^{(t_1-1)}(x_e) & \textit{if } b_{\mathrm{dl},f_i,e}^{(t_1-1,t_2-1)}(x_e) = b_{\mathrm{dl},f_j,e}^{(t_1-1,t_2-1)}(x_e) = 0 \\ \frac{1}{2}\lambda_{\mathrm{dl},e}^{(t_1)\prime}(x_e) & \textit{Otherwise} \end{cases}, \qquad x_e\in\mathcal{X}_e,\ e=(f_i,f_j)\in\mathcal{E},\ i<j, \tag{46}$$

   where

$$\lambda_{\mathrm{dl},e}^{(t_1)\prime}(x_e) \triangleq \log\left(\sum_{\boldsymbol{z}_{f_j}:\, z_e=x_e} f_j(\boldsymbol{z}_{f_j}) \cdot \prod_{e'\in\partial f_j\setminus\{e\}} \mu_{e'\to f_j}^{(t_1-1,t_2-1)}(z_{e'})\right) - \log\left(\sum_{\boldsymbol{z}_{f_i}:\, z_e=x_e} f_i(\boldsymbol{z}_{f_i}) \cdot \prod_{e'\in\partial f_i\setminus\{e\}} \mu_{e'\to f_i}^{(t_1-1,t_2-1)}(z_{e'})\right) + \log Z_{\mathrm{dl},f_i}^{(t_1-1,t_2-1)} - \log Z_{\mathrm{dl},f_j}^{(t_1-1,t_2-1)}.$$

   b) *When some termination criterion is met, we stop updating $\boldsymbol{\lambda}_{\mathrm{dl}}^{(t_1)}$ and switch to the outer loop.*
8) *For each $t_2\in\mathbb{Z}_{>0}$ and $e=(f_i,f_j)\in\mathcal{E}$, we fix $\boldsymbol{\lambda}_{\mathrm{dl}}^{(t_1)}$ and update $\boldsymbol{\gamma}^{(t_2)}$ as follows.*
   a) *If both $Z_{\mathrm{dl},f_j}^{(t_1,t_2-1)}$ and $Z_{\mathrm{dl},f_i}^{(t_1,t_2-1)}$ are positive-valued, we update $\gamma_{\mathrm{dl},e}^{(t_2)}(x_e)$ such that*

$$2\sqrt{\gamma_{\mathrm{dl},e}^{(t_2)}(x_e)} \propto \frac{\exp\left(\lambda_{\mathrm{dl},e,f_i}^{(t_1)}(x_e)\right)}{Z_{\mathrm{dl},f_i}^{(t_1,t_2-1)}} \cdot \sum_{\boldsymbol{z}_{f_i}:\, z_e=x_e} f_i(\boldsymbol{z}_{f_i}) \cdot \prod_{e'\in\partial f_i\setminus\{e\}} \mu_{e'\to f_i}^{(t_1,t_2-1)}(z_{e'}) + \frac{\exp\left(\lambda_{\mathrm{dl},e,f_j}^{(t_1)}(x_e)\right)}{Z_{\mathrm{dl},f_j}^{(t_1,t_2-1)}} \cdot \sum_{\boldsymbol{z}_{f_j}:\, z_e=x_e} f_j(\boldsymbol{x}_{f_j}) \cdot \prod_{e'\in\partial f_j\setminus\{e\}} \mu_{e'\to f_i}^{(t_1,t_2-1)}(z_{e'}), \qquad x_e\in\mathcal{X}_e, \tag{47}$$

$$\sum_{e} \gamma_{\mathrm{dl},e}^{(t_2)}(x_e) = 1, \qquad e=(f_i,f_j)\in\mathcal{E}. \tag{48}$$

9) *For each $t_2\in\mathbb{Z}_{>0}$, after updating $\boldsymbol{\gamma}_{\mathrm{dl}}^{(t_2)}$, we switch back to the inner loop, i.e., to step 7, to update $\boldsymbol{\lambda}_{\mathrm{dl}}^{(t_1)}$.*
10) *The outer loop, i.e., the update of $\boldsymbol{\gamma}_{\mathrm{dl}}^{(t_2)}$, is stopped when some termination criterion is met.*
11) *A collection of $\boldsymbol{\lambda}_{\mathrm{dl}}^{(t_1)}$ and $\boldsymbol{\gamma}_{\mathrm{dl}}^{(t_2)}$ is a called a fixed point of Algorithm 1 if*

$$\boldsymbol{\lambda}_{\mathrm{dl}}^{(t_1)} = \boldsymbol{\lambda}_{\mathrm{dl}}^{(t_1-1)}, \qquad \boldsymbol{\gamma}_{\mathrm{dl}}^{(t_2)} = \boldsymbol{\gamma}_{\mathrm{dl}}^{(t_2-1)}. \tag{49}$$

Proposition 19. *Each fixed point of Algorithm 1 with $\boldsymbol{\gamma}_e\in\mathcal{B}_e^{>}$ for all $e\in\mathcal{E}$ corresponds to a stationary point of $F_{\mathrm{B,d,N}}^{\mathrm{alt}}$.*

Proof. See Appendix D. ∎

[5] In general, the termination criterion is a combination of numerical convergence and an upper bound on the number of iterations.

Algorithm 1 The double-loop algorithm in Definition 18

Set $t_1 = t_2 = 0$.

Generate $\lambda_{\mathrm{dl},e}^{(t_1)}(x_e)$ following the uniform distribution in $[-1,1]$ for all $x_e \in \mathcal{X}_e$ and $e \in \mathcal{E}$.

Generate $\gamma_{\mathrm{dl}}^{(t_2)}(x_e)$ following the uniform distribution in $(0,1]$ for all $x_e \in \mathcal{X}_e$ and $e \in \mathcal{E}$.

repeat

    repeat

        for $n_e = 1$ to $|\mathcal{E}|$ **do**

            Update $\lambda_{\mathrm{dl},e}^{(t_1)}(x_e)$ according to (46).

            Increase t_1 by one.

        end for

    until Some termination criterion is met. (The end of inner loop)

    for $n_e = 1$ to $|\mathcal{E}|$ **do**

        Update $\gamma_{\mathrm{dl},e}^{(t_2)}(x_e)$ according to (47) and (48).

        Increase t_2 by one.

    end for

until Some termination criterion is met. (The end of outer loop)

Algorithm 2 The alternative SPA

Set $t_1 = t_2 = 0$.

Generate $\lambda_{\mathrm{dl},e}^{(t_1)}(x_e)$ following the uniform distribution in $[-1,1]$ for all $x_e \in \mathcal{X}_e$ and $e \in \mathcal{E}$.

Generate $\gamma_{\mathrm{dl}}^{(t_2)}(x_e)$ following the uniform distribution in $(0,1]$ for all $x_e \in \mathcal{X}_e$ and $e \in \mathcal{E}$.

repeat

    for $n_e = 1$ to $|\mathcal{E}|$ **do**

        Update $\lambda_{\mathrm{dl},e}^{(t_1)}(x_e)$ according to (46).

        Update $\gamma_{\mathrm{dl},e}^{(t_2)}(x_e)$ according to (50) and (51).

        Increase t_1 and t_2 by one, respectively.

    end for

until Some termination criterion is met.

Definition 20. *We give the details of Algorithm 2 in the following. There are two main differences between Algorithm 2 and Algorithm 1.*

1) *In Algorithm 2, we replace step 8a in Definition 18 by the following step.*
   - *If both $Z_{\mathrm{dl},f_j}^{(t_1,t_2-1)}$ and $Z_{\mathrm{dl},f_i}^{(t_1,t_2-1)}$ are positive-valued, we update $\gamma_{\mathrm{dl},e}^{(t_2)}(x_e)$ such that*

$$\sqrt{\gamma_{\mathrm{dl},e}^{(t_2)}(x_e)} \propto \left(\frac{\exp\left(\lambda_{\mathrm{dl},e,f_i}^{(t_1)}(x_e)\right)}{Z_{\mathrm{dl},f_i}^{(t_1,t_2-1)}} \cdot \left(\sum_{\boldsymbol{z}_{f_i}:\, z_e = x_e} f_i(\boldsymbol{z}_{f_i}) \cdot \prod_{e' \in \partial f_i \setminus \{e\}} \mu_{e' \to f_i}^{(t_1,t_2-1)}(z_{e'}) \right) \right.$$
$$\left. \cdot \frac{\exp\left(\lambda_{\mathrm{dl},e,f_j}^{(t_1)}(x_e)\right)}{Z_{\mathrm{dl},f_j}^{(t_1,t_2-1)}} \cdot \left(\sum_{\boldsymbol{z}_{f_j}:\, z_e = x_e} f_j(\boldsymbol{x}_{f_j}) \cdot \prod_{e' \in \partial f_j \setminus \{e\}} \mu_{e' \to f_i}^{(t_1,t_2-1)}(z_{e'}) \right) \right)^{1/2}, \tag{50}$$

$$\sum_{e} \exp\left(\lambda_{\mathrm{dl},e}^{(t_1)}(x_e)\right) \cdot \sqrt{\gamma_{\mathrm{dl},e}^{(t_2)}(x_e)} = 1. \tag{51}$$

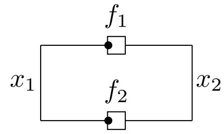


Fig. 3: S-NFG N_1.

2) *In Algorithm 2, we set $t_1 = t_2$ and we first update $\lambda_{\mathrm{dl},e}^{(t_1)}(x_e)$ according to steps 7a an 7b in Definition 18 and then update $\gamma_{\mathrm{dl},e}^{(t_2)}(x_e)$ following the previous step. After updating $\lambda_{\mathrm{dl},e}^{(t_2)}(x_e)$ and $\gamma_{\mathrm{dl},e}^{(t_2)}(x_e)$ for all $x_e \in \mathcal{E}$, we increase both t_1 and t_2 by one.*

Proposition 21. *Algorithm 2 is equivalent to the SPA in Definition 5. Each fixed point of Algorithm 2 with $\boldsymbol{\gamma}_e \in \mathcal{B}_e^{>}$ for all $e \in \mathcal{E}$ satisfies* (167) *and* (168)*, i.e., corresponds to a stationary point of $F_{\mathrm{B,d},\mathsf{N}}^{\mathrm{alt}}$.*

Proof. See Appendix E. ■

We conclude this section with some remarks.

1) Based on the definition of an S-NFG and its associated partition function and SPA in Section II, we have defined the Bethe partition function, which can be viewed as an approximation of the partition of the S-NFG. In Theorem 10, we have recalled the mains results of [9], which relate some of the SPA fixed points to some of the stationary point of the Bethe free energy function.
2) We have derived one of the dual formulations of the Bethe partition function of the S-NFG, which is denoted by $Z_{\mathrm{B,d},\mathsf{N}}^{*}$. We have also defined the quantity $Z_{\mathrm{B,d},\mathsf{N}}^{\mathrm{alt},*}$ which provides an upper bound of the Bethe partition function. We want to show that $Z_{\mathrm{B,d},\mathsf{N}}^{*} = Z_{\mathrm{B,d},\mathsf{N}}^{\mathrm{alt},*}$ in the remaining part of this paper. We have defined the sequences based on the locations of optimal value for the optimization problem in (21).
3) We have presented an algorithm for solving $Z_{\mathrm{B,d},\mathsf{N}}^{\mathrm{alt},*}$, which contains two loops. We have shown that the SPA can be recovered from this double-loop algorithm.

IV. The Analysis of a Single-Cycle S-NFG Example

In this section, in order to have a better understanding of our main results that will be presented in Section VII, we consider a simple single-cycle S-NFG N_1, as shown in Fig. 3. We want to relate one of the locations of the optimal value $F_{\mathrm{B,p},\mathsf{N}_1}^{*}$ for the optimization problem in (13) to an SPA fixed point for N_1. We propose two ways to relate these two concepts.

1) [9, Argument of Conjecture 2] For N_1 consisting of local functions f_1 and f_2, which equal zero for some x_1 and x_2, we consider a modification of the local functions by adding an infinitesimal positive-valued term to each zero factor. In this modified S-NFG, the location of the optimal value for the optimization problem in (13) corresponds to an SPA fixed point [9, Theorem 3]. In Item 1 of Theorem 27, we will show that when we let these infinitesimal positive-valued terms converge to zero, the SPA fixed point for the modified S-NFG converges to an SPA fixed point for the original S-NFG N_1 where some local functions' values are zero. Besides that, in (69) and (70) in Proposition 27, we will prove that the belief on each edge obtained by the SPA fixed point for the modified S-NFG converges to $1/2 \cdot \begin{pmatrix} 1, & 1 \end{pmatrix}^{\mathsf{T}}$ as the infinitesimal positive-valued terms go to zero.
2) Another way is that we consider the dual formulation of the optimization problem in (13), i.e., the optimization problem in (21) with optimal value $Z_{\mathrm{B,d},\mathsf{N}_1}^{\mathrm{alt},*}$. We let $\boldsymbol{\gamma}^{(m)}$ and $\boldsymbol{\lambda}^{(n)}$ be sequences such that the associated objective function $Z_{\mathrm{B,d},\mathsf{N}_1}(\boldsymbol{\gamma}^{(m)}, \boldsymbol{\lambda}^{(n)})$,

as defined in (18), converges to $Z^{\mathrm{alt},*}_{\mathrm{B,d,N_1}}$.[6] Then in Theorem 40, we will prove that the collection of messages $\boldsymbol{\mu}^{(m,n)}$ converges to an SPA fixed point of N_1, where $\boldsymbol{\mu}^{(m,n)}$ is given by

$$\boldsymbol{\mu}^{(m,n)} = \bigl(\mu^{(m,n)}_{e\to f}(x_e)\bigr)_{x_e\in\mathcal{X}_e,\, e\in\partial f,\, f\in\mathcal{F}},$$
$$\mu^{(m,n)}_{e\to f}(x_e) \triangleq \exp\bigl(\lambda^{(n)}_{e,f}(x_e)\bigr)\cdot\sqrt{\gamma^{(m)}_e(x_e)}.$$

Also, in Theorem 36, we will prove that the set of beliefs $\boldsymbol{b}^{(m,n)}_{\mathcal{F}}$ defined in (38), which are functions of $\boldsymbol{\gamma}^{(m)}$ and $\boldsymbol{\lambda}^{(n)}$, satisfies

$$\lim_{n\to\infty} b^{(m,n)}_{f_1,e}(x_e) = \lim_{n\to\infty} b^{(m,n)}_{f_2,e}(x_e) = \gamma^{(m)}_e(x_e), \qquad x_e\in\mathcal{X}_e, \quad m\in\mathbb{Z}_{>0}, \quad \boldsymbol{\gamma}^{(m)}_e\in\mathcal{B}^{>}_e,\, e\in\mathcal{E}, \tag{52}$$

which shows that by varying $\boldsymbol{\gamma}^{(m)}$, we can let the marginals of the beliefs on edges $\boldsymbol{b}^{(m,n)}_{\mathcal{F}}$ converge to any point in $\prod_e \mathcal{B}^{>}_e$. As we will see in Theorems 36 and 40, by the second method, we can show that any point in $\mathcal{B}(\mathsf{N}_1)$ with

$$\boldsymbol{b}_{f_1} = \boldsymbol{b}_{f_2} = \begin{pmatrix} \gamma^{(m)}_1(1) & 0 \\ 0 & \gamma^{(m)}_1(2) \end{pmatrix}, \qquad \boldsymbol{\gamma}^{(m)}_1 \in \mathcal{B}^{>}_1,$$

corresponds an SPA fixed point of N_1, which indicates that the second method is a promising method for generalizing our results in this section to the general S-NFGs.

Let us provide some technical details of the S-NFG N_1 first.

Definition 22. *We make the following definitions for* N_1.

1) *The dots in Fig. 3 are used for denoting the row indices of matrices* $\boldsymbol{f}_1$ *and* $\boldsymbol{f}_2$ *associated with function nodes* f_1 *and* f_2, *respectively. To be more specific, the rows in the following matrices*

$$\boldsymbol{f}_1 \triangleq \Bigl(f_1(x_1,x_2)\Bigr)_{x_1,x_2}, \qquad \boldsymbol{f}_2 \triangleq \Bigl(f_2(x_1,x_2)\Bigr)_{x_1,x_2}$$

are indexed by x_1 *and the columns are indexed by* x_2. *Because we want to discuss different* $\boldsymbol{f}_1$ *and* $\boldsymbol{f}_2$ *in* N_1, *the details of* $\boldsymbol{f}_1$ *and* $\boldsymbol{f}_2$ *will be given in the coming sections.*

2) *The set of edges is given by* $\mathcal{E} = [2]$.

3) *The variables* x_1 *and* x_2 *take value in the alphabet* $\mathcal{X}_1 = \mathcal{X}_2 = [2] = \{1,2\}$.

Based on the previous definitions, we further investigate the function $F^{\mathrm{alt},*}_{\mathrm{B,d,N_1}}$ and the associated sequences $\boldsymbol{\gamma}^{(m)}$ and $\boldsymbol{\lambda}^{(n)}$ for N_1, which are specified in Definition 17.

Remark 23. *Let us provide specific properties of the dual function* $F^{\mathrm{alt},*}_{\mathrm{B,d,N_1}}$ *and the associated sequences* $\boldsymbol{\gamma}^{(m)}$ *and* $\boldsymbol{\lambda}^{(n)}$ *for* N_1.

1) *For simplicity, we do not consider the sequence* $\{\alpha^{(k)}\}_k$ *in this example because it is unnecessary for obtaining the main results in this section.*

2) *By the definition of* $\boldsymbol{\gamma}^{(m)}$ *in* (29), *the collection of vectors* $\boldsymbol{\gamma}^{(m)}$ *satisfies*

$$\gamma^{(m)}_1(1) + \gamma^{(m)}_1(2) = 1, \quad \gamma^{(m)}_2(1) + \gamma^{(m)}_2(2) = 1, \quad m\in\mathbb{Z}_{>0}. \tag{53}$$

3) *By the definition of* $F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma},\boldsymbol{\lambda})$ *in* (17) *for* N_1, *the function* $F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma}^{(m)},\boldsymbol{\lambda}^{(n)})$ *equals*

$$F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma}^{(m)},\boldsymbol{\lambda}^{(n)}) \overset{(a)}{=} -\log Z_{f_1}\bigl(\boldsymbol{\gamma}^{(m)},\,\boldsymbol{\lambda}^{(n)}\bigr) - \log Z_{f_2}\bigl(\boldsymbol{\gamma}^{(m)},\,\boldsymbol{\lambda}^{(n)}\bigr). \tag{54}$$

[6]Although by Propositions 15 and 16, we only know that $Z^{\mathrm{alt},*}_{\mathrm{B,d,N_1}} \geq Z^{*}_{\mathrm{B,d,N_1}} = Z^{*}_{\mathrm{B,p,N_1}}$, in Theorem 57 in Section VII, we will prove that $Z^{\mathrm{alt},*}_{\mathrm{B,d,N}} = Z^{*}_{\mathrm{B,p,N}}$, which shows that the optimization problem in (21) is the dual formulation of the Bethe partition function.

4) *The functions* $b_{f_1}^{(m,n)}(x_1,x_2)$ *and* $b_{f_2}^{(m,n)}(x_1,x_2)$ *defined in* (37) *can be written as*

$$b_{f_1}^{(m,n)}(x_1,x_2) = f_1(x_1,x_2) \cdot \frac{\exp\Big(\lambda_1^{(n)}(x_1) + \lambda_2^{(n)}(x_2)\Big) \cdot \sqrt{\gamma_1^{(m)}(x_1) \cdot \gamma_2^{(m)}(x_2)}}{Z_{f_1}\big(\boldsymbol{\gamma}^{(m)},\, \boldsymbol{\lambda}^{(n)}\big)}, \qquad Z_{f_1}\big(\boldsymbol{\gamma}^{(m)},\, \boldsymbol{\lambda}^{(n)}\big) > 0, \tag{55}$$

$$b_{f_2}^{(m,n)}(x_1,x_2) = f_2(x_1,x_2) \cdot \frac{\exp\Big(-\lambda_1^{(n)}(x_1) - \lambda_2^{(n)}(x_2)\Big) \cdot \sqrt{\gamma_1^{(m)}(x_1) \cdot \gamma_2^{(m)}(x_2)}}{Z_{f_2}\big(\boldsymbol{\gamma}^{(m)},\, \boldsymbol{\lambda}^{(n)}\big)}, \qquad Z_{f_2}\big(\boldsymbol{\gamma}^{(m)},\, \boldsymbol{\lambda}^{(n)}\big) > 0. \tag{56}$$

We also define the matrices $\boldsymbol{b}_{f_1}^{(m,n)}$ *and* $\boldsymbol{b}_{f_2}^{(m,n)}$ *to be*

$$\boldsymbol{b}_{f_i}^{(m,n)} \triangleq \begin{pmatrix} b_{f_i}^{(m,n)}(1,1) & b_{f_i}^{(m,n)}(1,2) \\ b_{f_i}^{(m,n)}(2,1) & b_{f_i}^{(m,n)}(1,2) \end{pmatrix}, \qquad i \in [2]. \tag{57}$$

The collection of belief $\boldsymbol{b}_{\mathcal{F}}^{(m,n)}$ *is defined to be*

$$\boldsymbol{b}_{\mathcal{F}}^{(m,n)} \triangleq \Big(\boldsymbol{b}_{f_1}^{(m,n)} \quad \boldsymbol{b}_{f_2}^{(m,n)}\Big).$$

5) *The marginal* $b_{f_i,e}^{(m,n)}$ *given in* (25) *can be rewritten as*

$$b_{f_i,1}^{(m,n)}(x_1) = \sum_{x_2} b_{f_i}^{(m,n)}(x_1,x_2), \qquad b_{f_i,2}^{(m,n)}(x_2) = \sum_{x_1} b_{f_i}^{(m,n)}(x_1,x_2), \qquad i \in [2]. \tag{58}$$

For the S-NFG N_1 given in Definition 22, we can obtain explicit expressions for the sequences $\{\boldsymbol{\gamma}^{(m)}\}$ and $\{\boldsymbol{\lambda}^{(n)}\}$. Although these expressions are not unique, analyzing the sequences with explicit expressions helps us to better understand the proof of the main results for this single-cycle S-NFG and even for the general S-NFGs.

A. An Example Single-Cycle S-NFG with Positive-Valued Entries Only

In this subsection, we discuss N_1 in Fig. 3 with positive-valued local functions only.

Example 24. *For* $r \in \mathbb{R}_{>0}$, *we consider* N_1, *where the local functions associated with function nodes* f_1 *and* f_2 *are given by* $\boldsymbol{f}_{1,r}$ *and* $\boldsymbol{f}_{2,r}$*:*

$$\boldsymbol{f}_{1,r} = \begin{pmatrix} f_{1,r}(1,1) & f_{1,r}(1,2) \\ f_{1,r}(2,1) & f_{1,r}(2,2) \end{pmatrix} \triangleq \begin{pmatrix} 1 & 1 \\ \delta_1(r) & 1 \end{pmatrix}, \tag{59}$$

$$\boldsymbol{f}_{2,r} = \begin{pmatrix} f_{2,r}(1,1) & f_{2,r}(1,2) \\ f_{2,r}(2,1) & f_{2,r}(2,2) \end{pmatrix} \triangleq \begin{pmatrix} 1 & \delta_2(r) \\ \delta_3(r) & 1 \end{pmatrix}, \tag{60}$$

which means that

$$\boldsymbol{f}_1 = \boldsymbol{f}_{1,r}, \qquad \boldsymbol{f}_2 = \boldsymbol{f}_{2,r}.$$

Here, the functions $\delta_1(r)$, $\delta_2(r)$, *and* $\delta_3(r)$ *are arbitrary functions such that*

$$\delta_1(r),\, \delta_2(r),\, \delta_3(r) \in \mathbb{R}_{>0}, \qquad r \in \mathbb{R}_{>0}, \tag{61}$$

$$\lim_{r \downarrow 0} \delta_i(r) = 0, \qquad i \in [3]. \tag{62}$$

In this section, we let r go to zero and thus some of the entries in the local functions converge to zero. We want to study whether the associated location of the optimal value for the primal formulation converges. Let us make some definitions for the above-considered S-NFG N_1.

Definition 25. *The vector* $\boldsymbol{\delta}(r)$ *is defined to be*

$$\boldsymbol{\delta}(r) \triangleq \Big(\delta_1(r) \quad \delta_2(r) \quad \delta_3(r)\Big). \tag{63}$$

The matrix $\boldsymbol{G}_{\mathsf{N}_1,r}$ is defined to be

$$\begin{aligned}\boldsymbol{G}_{\mathsf{N}_1,r} &\triangleq \boldsymbol{f}_{1,r} \cdot \boldsymbol{f}_{2,r}^{\mathsf{T}} \\ &= \begin{pmatrix} \delta_2(r)+1 & \delta_3(r)+1 \\ \delta_1(r)+\delta_2(r) & \delta_1(r)\cdot\delta_3(r)+1 \end{pmatrix} \in \mathbb{R}_{>0}^{2\times 2}. \end{aligned} \tag{64}$$

Let $\Lambda_{\max}(r)$ be the eigenvalue of $\boldsymbol{G}_{\mathsf{N}_1,r}$ with the largest magnitude and $\boldsymbol{v}_{\mathrm{L}}$ and $\boldsymbol{v}_{\mathrm{R}}$ be the associated left and right eigenvectors, respectively. It follows from Perron-Frobenius theory that $\Lambda_{\max}(r)$ is unique and nonnegative real-valued, and that $\boldsymbol{v}_{\mathrm{L}}$ and $\boldsymbol{v}_{\mathrm{R}}$ can be chosen to have only positive entries. (See, e.g., [18, Section 8.3].) For convenience, we define $c_1(r)$ to be

$$c_1(r) \triangleq \sqrt{\big(\delta_1(r)\cdot\delta_3(r)-\delta_2(r)\big)^2 + 4\big(\delta_1(r)+\delta_2(r)\big)\cdot\big(1+\delta_3(r)\big)}.$$

Proposition 26. *The following properties hold for the S-NFG N_1 considered in Definition 22 and Example 24.*

1) *[16, Corollary 2] Both the Bethe free energy function $F_{\mathrm{B,p},\mathsf{N}_1}$ and its alternative $F_{\mathrm{B,p},\mathsf{N}_1}^{(1)}$ are convex w.r.t. $\boldsymbol{\beta} \in \mathcal{B}(\mathsf{N}_1)$, where $F_{\mathrm{B,p},\mathsf{N}_1}$ and $F_{\mathrm{B,p},\mathsf{N}_1}^{(1)}$ for N_1 are defined in* (10) *and* (11)*, respectively, and the set $\mathcal{B}(\mathsf{N}_1)$ is defined in* (7).
2) *[9, Theorem 3] The location for the optimal value $F_{\mathrm{B,p},\mathsf{N}_1}^{*}$ are given by the SPA fixed-point messages.*
3) *If $|\delta_i(r)| < 1$ for all $i \in [3]$, we have*

$$c_1(r) \geq \max\Big(2\delta_2(r),\, 2\delta_1(r)\cdot\delta_3(r)\Big). \tag{65}$$

4) *The SPA fixed-point messages $\boldsymbol{\mu}_{1\to f_1}^{(t)}$ and $\boldsymbol{\mu}_{1\to f_2}^{(t)}$ are proportional to $\boldsymbol{v}_{\mathrm{L}}$ and $\boldsymbol{v}_{\mathrm{R}}$, respectively, i.e., $\boldsymbol{\mu}_{1\to f_1}^{(t)} \propto \boldsymbol{v}_{\mathrm{L}}$ and $\boldsymbol{\mu}_{1\to f_2}^{(t)} \propto \boldsymbol{v}_{\mathrm{R}}$.*
5) *The beliefs of function nodes evaluated at the SPA fixed point, as defined in* (2)*, satisfy*

$$\beta_{f_1}(x_1,x_2) \propto v_{\mathrm{L}}(x_1)\cdot f_{1,r}(x_1,x_2)\left(\sum_{x_1'} \cdot f_{2,r}(x_1',x_2)\cdot v_{\mathrm{R}}(x_1')\right), \tag{66}$$

$$\beta_{f_2}(x_1,x_2) \propto \left(\sum_{x_1'} v_{\mathrm{L}}(x_1')\cdot f_{1,r}(x_1',x_2)\right)\cdot f_{2,r}(x_1,x_2)\cdot v_{\mathrm{R}}(x_1). \tag{67}$$

6) *The associated Bethe partition function $Z_{\mathrm{B,p},\mathsf{N}_1}^{*}$ satisfies*

$$Z_{\mathrm{B,p},\mathsf{N}_1}^{*} = Z_{\mathrm{B,d},\mathsf{N}_1}^{*} = \Lambda_{\max}(r) = F_{\mathrm{B,p},\mathsf{N}_1}(\boldsymbol{\beta}) \qquad \text{s.t. } \boldsymbol{\beta}\in\mathcal{B}(\mathsf{N}_1), \text{ where } \boldsymbol{\beta}_{f_1} \text{ satisfies (66) and } \boldsymbol{\beta}_{f_2} \text{ satisfies (67)}. \tag{68}$$

Proof. See Appendix F. ■

Then we analyze the properties of the beliefs at the SPA fixed point and the primal formulation as $r \downarrow 0$.

Proposition 27. *The following properties hold for the S-NFG N_1 considered in Example 24.*

1) *The eigenvalue $\Lambda_{\max}(r)$ for matrix $\boldsymbol{G}_{\mathsf{N}_1,r}$, as specified in Definition 25, satisfy*

$$\lim_{r\downarrow 0}\Lambda_{\max}(r) = 1.$$

The associated eigenvectors, as specified in Definition 25, satisfy

$$\lim_{r\downarrow 0}\boldsymbol{v}_{\mathrm{L}}(r) = \begin{pmatrix}0 & 1\end{pmatrix}^{\mathsf{T}}, \qquad \lim_{r\downarrow 0}\boldsymbol{v}_{\mathrm{R}}(r) = \begin{pmatrix}1 & 0\end{pmatrix}^{\mathsf{T}}.$$

2) *The SPA fixed-point messages $\boldsymbol{\mu}^{(t)}$ satisfy*

$$\lim_{r\downarrow 0}\mu_{1\to f_1}^{(t)} = \begin{pmatrix}0 & 1\end{pmatrix}^{\mathsf{T}}, \qquad \lim_{r\downarrow 0}\mu_{1\to f_2}^{(t)} = \begin{pmatrix}1 & 0\end{pmatrix}^{\mathsf{T}},$$

$$\lim_{r\downarrow 0}\mu_{2\to f_1}^{(t)} = \begin{pmatrix}1 & 0\end{pmatrix}^{\mathsf{T}}, \qquad \lim_{r\downarrow 0}\mu_{2\to f_2}^{(t)} = \begin{pmatrix}0 & 1\end{pmatrix}^{\mathsf{T}}.$$

3) *The beliefs evaluated at the SPA fixed point satisfy*

$$\lim_{r \downarrow 0} \boldsymbol{\beta}_{f_1} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{69}$$

$$\lim_{r \downarrow 0} \boldsymbol{\beta}_{f_2} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{70}$$

Proof. See Appendix G. ■

For comparison, we also study the dual formulation.

Proposition 28. *The following properties hold for the S-NFG N_1 considered in Example 24.*

1) *If we set $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ to be the collections of the vectors satisfying*

$$\gamma_e(x_e) \propto \sqrt{\mu_{e,f_1}^{(t)}(x_e) \cdot \mu_{e,f_2}^{(t)}(x_e)}, \qquad x_e \in \mathcal{X}_e,\, \boldsymbol{\gamma}_e \in \mathcal{B}_e^{>},\, e \in \mathcal{E}, \tag{71}$$

$$\exp(\lambda_e(x_e)) = \sqrt{\frac{\mu_{e,f_1}^{(t)}(x_e)}{\mu_{e,f_2}^{(t)}(x_e)}}, \qquad x_e \in \mathcal{X}_e,\, e \in \mathcal{E}, \tag{72}$$

where the SPA fixed-point messages $\boldsymbol{\mu}^{(t)}$ are given in (174)–(177) *in the proof of Proposition 27, then the above given $(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ are at the location of the optimal value for the dual formulation in* (19)*. If we let $r \downarrow 0$, the associated optimal location of the optimal value satisfies*

$$\lim_{r \downarrow 0} \boldsymbol{\gamma}_e = \frac{1}{2} \begin{pmatrix} 1 & 1 \end{pmatrix}^{\mathsf{T}}, \qquad e \in \mathcal{E}, \tag{73}$$

$$\lim_{r \downarrow 0} \begin{pmatrix} \exp(\lambda_1(1)) & \exp(\lambda_1(2)) \end{pmatrix} = \begin{pmatrix} 0 & \infty \end{pmatrix}, \tag{74}$$

$$\lim_{r \downarrow 0} \begin{pmatrix} \exp(\lambda_2(1)) & \exp(\lambda_2(2)) \end{pmatrix} = \begin{pmatrix} \infty & 0 \end{pmatrix}. \tag{75}$$

2) *The quantity $Z_{\mathrm{B,d},\mathsf{N}_1}^{\mathrm{alt},*}$ equals the Bethe partition function of*

$$Z_{\mathrm{B,d},\mathsf{N}_1}^{\mathrm{alt},*} = Z_{\mathrm{B,d},\mathsf{N}_1}^{*} = Z_{\mathrm{B,p},\mathsf{N}_1}^{*} = \Lambda_{\max}. \tag{76}$$

Proof. See Appendix H. ■

We summarize the above results as follows.

1) Because all the local functions are positive-valued, in Proposition 26, we showed that the locations of the optimal values for both primal and dual formulations correspond to an SPA fixed point.
2) Because N_1 in Fig. 3 is a single-cycle S-NFG, the SPA is equivalent to applying the power method on the matrix associated with N_1, and the SPA fixed-point messages correspond to the eigenvectors of the matrices $\boldsymbol{f}_{1,r} \cdot \boldsymbol{f}_{2,r}^{\mathsf{T}}$ and $\boldsymbol{f}_{1,r}^{\mathsf{T}} \cdot \boldsymbol{f}_{2,r}$.
3) In (69) and (70) in Proposition 28, we showed that the location of the optimal value for the primal formulation converge as follows:

$$\lim_{r \downarrow 0} \beta_{f_1,1}(x_1) = \lim_{r \downarrow 0} \beta_{f_2,1}(x_1) = \lim_{r \downarrow 0} \beta_{f_1,2}(x_1) = \lim_{r \downarrow 0} \beta_{f_2,2}(x_1) = \frac{1}{2}, \qquad x_1 \in \mathcal{X}_e.$$

4) In Proposition 28, we showed that one of the locations of the optimal value for the dual formulation satisfies

$$\lim_{r \downarrow 0} \boldsymbol{\gamma}_e = \frac{1}{2} \begin{pmatrix} 1 & 1 \end{pmatrix}^{\mathsf{T}}, \qquad e \in \mathcal{E}. \tag{77}$$

B. An Example Single-Cycle S-NFG with Zero Entries

In this subsection, we consider N_1 in Fig. 3 with the details specified in Example 29. The remaining details of N_1 are given in Definition 22. In this S-NFG, we consider local functions $f_{1,r}^{(1)}$ and f_2 for function nodes f_1 and f_2 such that $\lim_{r\downarrow 0} f_{1,r}^{(1)} = f_1$. (For details, see Example 29 below.) We want to see that whether the beliefs evaluated at the associated SPA fixed point still converge to the limits in (69) and (70) as r goes to zero.

Example 29. *In this section, we consider S-NFG N_1, where the local functions for function nodes f_1 and f_2 are given by*

$$\boldsymbol{f}_{1,r}^{(1)} = \begin{pmatrix} f_{1,r}^{(1)}(1,1) & f_{1,r}^{(1)}(1,2) \\ f_{1,r}^{(1)}(2,1) & f_{1,r}^{(1)}(2,2) \end{pmatrix} \triangleq \begin{pmatrix} 1+\delta_2(r) & 1 \\ \delta_1(r) & 1 \end{pmatrix}, \qquad \boldsymbol{f}_2 = \begin{pmatrix} f_2(1,1) & f_2(1,2) \\ f_2(2,1) & f_2(2,2) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

where $\delta_1(r)$ and $\delta_2(r)$ are defined in Example 24. Let $\Lambda_{\max}^{(1)}(r)$ be the eigenvalue of the matrix $\boldsymbol{f}_{1,r}^{(1)}$ with the largest magnitude and $\boldsymbol{v}_{\mathrm{L}}^{(1)}$ and $\boldsymbol{v}_{\mathrm{R}}^{(1)}$ be the associated left and right eigenvectors, respectively.

Proposition 26 still holds for the S-NFG N_1 considered in Definition 22 and Example 29 (modulo the notational changes mentioned in the following proposition).

Proposition 30. *Proposition 26 still holds for the S-NFG N_1 considered in Definition 22 and Example 29 after making the following the notational changes.*

1) *The local functions $f_{1,r}$ and $f_{2,r}$ are replaced by $f_{1,r}^{(1)}$ and f_2, respectively.*
2) *The eigenvalue $\Lambda_{\max}(r)$ is replaced by $\Lambda_{\max}^{(1)}(r)$.*
3) *The vectors $\boldsymbol{v}_{\mathrm{L}}$ and $\boldsymbol{v}_{\mathrm{R}}$ are replaced by $\boldsymbol{v}_{\mathrm{L}}^{(1)}$ and $\boldsymbol{v}_{\mathrm{R}}^{(1)}$, respectively*

Proof. The proof is similar to the proof of Proposition 26 and thus it is omitted here. ■

However, some of the properties in Proposition 27 do not hold for the S-NFG N_1 considered in this section.

Proposition 31. *The following properties hold for the S-NFG N_1 considered in Definition 22 and Example 29.*

1) *Items 1 and 2 in Proposition 27 hold after making the the notational changes in Proposition 30.*
2) *If we set $\delta_2(r) = \sqrt{c_\delta \cdot \delta_1(r)}$ for some $c_\delta \in \mathbb{R}_{\geq 0}$, the beliefs evaluated at the associated SPA fixed point satisfy*

$$\lim_{r\downarrow 0} \boldsymbol{\beta}_{f_1} = \begin{pmatrix} \frac{1}{2} + \frac{f_\delta(c_\delta)}{8+2f_\delta(c_\delta)} & 0 \\ 0 & \frac{1}{2} - \frac{f_\delta(c_\delta)}{8+2f_\delta(c_\delta)} \end{pmatrix}, \tag{78}$$

$$\lim_{r\downarrow 0} \boldsymbol{\beta}_{f_2} = \lim_{r\downarrow 0} \boldsymbol{\beta}_{f_1}, \tag{79}$$

where $f_\delta(c_\delta) \triangleq c_\delta + \sqrt{c_\delta^2 + 4c_\delta} \in \mathbb{R}_{>0}$. For any $1/2 \leq \alpha_\delta < 1$, we can find a $c_\delta \in \mathbb{R}_{>0}$ such that

$$\lim_{r\downarrow 0} \boldsymbol{\beta}_{f_1} = \lim_{r\downarrow 0} \boldsymbol{\beta}_{f_2} = \begin{pmatrix} \alpha_\delta & 0 \\ 0 & 1-\alpha_\delta \end{pmatrix}.$$

Proof. The proof is similar to the proof of Proposition 27 and thus it is omitted here. ■

Comparing Proposition 31 with Proposition 27, for the S-NFG considered in this paper, we can let the beliefs evaluated at the SPA fixed point converge to different limits instead of the limits in (69) and (70) by suitably defining $\delta_2(r)$ based on $\delta_1(r)$.

C. Another Example Single-Cycle S-NFG with Zero Entries

In this subsection, we consider N_1 in Fig. 3 with the details specified in the upcoming Example 32. The remaining details of N_1 are as in Definition 22. There are two main goals in the subsection.

1) The first goal is to compare the results obtained in this subsection with the results obtained in Section IV-A.
   a) In Section IV-A, we considered a modified S-NFG such that all the local function are positive-valued, and we found the locations of the optimal value for the primal and dual formulations. As proven in Propositions 26 and 28, these locations correspond to an SPA fixed point. Then we some entries in the local functions converge to zero. The associated locations and the SPA fixed-point messages converge. In particular, the SPA fixed point for the modified S-NFG converges to the SPA fixed point of the original S-NFG with zero-valued local functions.
   b) In this section, we directly consider the S-NFG with zero-valued local functions, and we will show that any collection of vectors $\boldsymbol{\gamma} \in \prod_e \mathcal{B}_e^{>}$ with $\gamma_1 = \gamma_2$ corresponds to the location of the optimal value $F_{\mathrm{B,d,N_1}}^{\mathrm{alt},*}$ of the optimization problem in (22) as well as an SPA fixed point of the S-NFG N_1. (For details, see Proposition 35 and Theorem 40.) It is different from the results obtained in Section IV-A, where we only show that the $\boldsymbol{\gamma}$ in (77) corresponds to an SPA fixed point of N_1 considered in Example 32.
2) The second goal is to understand the main idea of the proof for the general S-NFG in Section VII by analyzing this simple example. Note that the main results in this subsection are obtained by analyzing the location of the optimal value for the optimization problem in (22), which is similar to the main idea in Section VII. In Definitions 33, we will specify the sequence $\{\boldsymbol{\lambda}^{(n)}\}_{n\in\mathbb{Z}_{>0}}$ and the vector $\boldsymbol{\gamma}^{(1)}$ such that $F_{\mathrm{B,d,N_1}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\lambda}^{(n)})$ converge to $F_{\mathrm{B,d,N_1}}^{\mathrm{alt},*}$ as $n \to \infty$. Based on that, in Theorem 38, we will show $F_{\mathrm{B,d,N_1}}^{\mathrm{alt},*} = F_{\mathrm{B,d,N_1}}^{*} = F_{\mathrm{B,p,N_1}}^{*}$, and in Theorem 40, we will prove that the collection of messages $\boldsymbol{\mu}^{(1,n)}$ satisfying (91) converges to SPA fixed-point messages. The idea of these Theorems' proof is similar to the idea of the main results' proof in Section VII.

Example 32. *The local functions for function node f_1 and f_2 are given by $\lim_{r\downarrow 0} \boldsymbol{f}_{1,r}$ and $\lim_{r\downarrow 0} \boldsymbol{f}_{2,r}$*

$$\boldsymbol{f}_1 = \lim_{r\downarrow 0} \boldsymbol{f}_{1,r} = \lim_{r\downarrow 0} \begin{pmatrix} 1 & 1 \\ \delta_1(r) & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{f}_2 = \lim_{r\downarrow 0} \boldsymbol{f}_{2,r} = \lim_{r\downarrow 0} \begin{pmatrix} 1 & \delta_2(r) \\ \delta_3(r) & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

where $\delta_1(r)$, $\delta_2(r)$, and $\delta_3(r)$ are defined in Example 24. Note that the left and right eigenvectors of $\boldsymbol{f}_1$ are $(0\ 1)^{\mathsf{T}}$ and $(1\ 0)^{\mathsf{T}}$, respectively.

We define the sequences $\{\boldsymbol{\gamma}^{(k)}\}_k$ and $\{\boldsymbol{\lambda}^{(n)}\}_n$ in the following such that they correspond to the locations of the optimal value for the dual formulation. (For details, see Proposition 35 in the following text.)

Definition 33. *We make the following definitions for the sequences $\{\boldsymbol{\gamma}^{(k)}\}_k$ and $\{\boldsymbol{\lambda}^{(n)}\}_n$.*

1) *We fix $k = 1$ and define $\boldsymbol{\gamma}^{(1)}$ to be*

$$\boldsymbol{\gamma}_e^{(1)} \in \mathcal{B}_e^{>}, \qquad e \in \mathcal{E} = [2],$$
$$\gamma_1^{(1)}(x_e) = \gamma_2^{(1)}(x_e), \qquad x_e \in \mathcal{X}_e. \tag{80}$$

2) *We consider the following sequence of $\{\boldsymbol{\lambda}^{(n)}\}_n$:*

$$\lambda_1^{(n)}(1) = -n, \quad \lambda_1^{(n)}(2) = 0, \quad \lambda_2^{(n)}(1) = n, \quad \lambda_2^{(n)}(2) = 0.$$

The definition of $\boldsymbol{\lambda}^{(n)}$ in Definition 33 implies

$$\exp\Big(\lambda_1^{(n)}(x_1) + \lambda_2^{(n)}(x_1)\Big) = 1, \qquad x_1 \in [2],\, n \in \mathbb{Z}_{>0},$$

$$\lim_{n\to\infty} \exp\Big(\lambda_1^{(n)}(1) + \lambda_2^{(n)}(2)\Big) = \lim_{n\to\infty} \exp(-n) = 0.$$

Remark 34. *We make the following remarks for the functions and the sequences relating to the S-NFG N_1 specified in Definition 22 and Example 32.*

1) *The associated Bethe free energy function $F_{\mathrm{B,p,N}}$ defined in* (10) *for N_1 satisfies*

$$F_{\mathrm{B,p,N_1}}(\boldsymbol{\beta}) = 0, \qquad \boldsymbol{\beta} \in \mathcal{B}(\mathsf{N}_1), \tag{81}$$

where $\mathcal{B}(\mathsf{N}_1)$ is defined in (7)*. The above expression implies*

$$F^{*}_{\mathrm{B,p,N_1}} = 0. \tag{82}$$

2) *The function Z_{f_1} defined in* (16) *equals*

$$\begin{aligned} Z_{f_1}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) &= \sum_{x_1,x_2 \in \{(1,1),(1,2),(2,2)\}} \exp\Big(\lambda_{1,f_1}(x_1) + \lambda_{2,f_1}(x_2)\Big) \cdot \sqrt{\gamma_1(x_1)\cdot\gamma_2(x_2)} \\ &\overset{(a)}{=} \exp\Big(\lambda_1(1) + \lambda_2(1)\Big) \cdot \sqrt{\gamma_1(1)\cdot\gamma_2(1)} \\ &\quad + \exp\Big(\lambda_1(1) + \lambda_2(2)\Big) \cdot \sqrt{\gamma_1(1)\cdot\gamma_2(2)} \\ &\quad + \exp\Big(\lambda_1(2) + \lambda_2(2)\Big) \cdot \sqrt{\gamma_1(2)\cdot\gamma_2(2)}, \end{aligned} \tag{83}$$

where step (a) follows from the definition of $\boldsymbol{\lambda}_{e,f}$ in (15)*. Similarly, the function Z_{f_2} defined in* (84) *equals*

$$\begin{aligned} Z_{f_2}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) = \exp\Big(-\big(\lambda_1(1) + \lambda_2(1)\big)\Big) \cdot \sqrt{\gamma_1(1)\cdot\gamma_2(1)} \\ + \exp\Big(-\big(\lambda_1(2) + \lambda_2(2)\big)\Big) \cdot \sqrt{\gamma_1(2)\cdot\gamma_2(2)}. \end{aligned} \tag{84}$$

3) *By the definition of $\boldsymbol{\gamma}^{(1)}$ in Definition 33, we have $\boldsymbol{\gamma}_e^{(1)} \in \mathcal{B}_e^{>}$ for all $e \in \mathcal{E} = [2]$. Then the functions Z_{f_1} and Z_{f_2} satisfy*

$$Z_{f_1}\big(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\lambda}^{(n)}\big),\, Z_{f_2}\big(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\lambda}^{(n)}\big) \in \mathbb{R}_{>0}, \qquad n \in \mathbb{Z}_{>0}, \tag{85}$$

which implies $b_{f_1}^{(1,n)}$ in (55) *and $b_{f_2}^{(1,n)}$ in* (56) *are well defined, i.e.,*

$$\begin{aligned} &\exists (x_1,x_2) \in [2]\times[2] \qquad \text{s.t. } \boldsymbol{b}_{f_1}^{(1,n)}(x_1,x_2) \in \mathbb{R}_{>0}, \qquad n \in \mathbb{Z}_{>0}, \\ &\exists (x_1',x_2') \in [2]\times[2] \qquad \text{s.t. } \boldsymbol{b}_{f_2}^{(1,n)}(x_1',x_2') \in \mathbb{R}_{>0}, \qquad n \in \mathbb{Z}_{>0}. \end{aligned}$$

We focus on the optimization problem defined in (22) first. We show that the considered sequence $\{\boldsymbol{\lambda}^{(n)}\}_{n\in\mathbb{Z}_{>0}}$ converges to the location of the optimal value $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}$ and the sequences $\{\boldsymbol{\lambda}^{(n)}\}_{n\in\mathbb{Z}_{>0}}$ and $\boldsymbol{\gamma}^{(1)}$ converges to one of the locations of the optimal value for the dual formulation.

Proposition 35. *Consider the S-NFG N_1 specified in Definition 22 and Example 32.*

1) *It holds that*

$$\begin{aligned} \lim_{n\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}^{(n)}) &= \hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma}) \\ &= -2\log\Big(\sqrt{\gamma_1(1)\cdot\gamma_2(1)} + \sqrt{\gamma_1(2)\cdot\gamma_2(2)}\Big), \qquad \boldsymbol{\gamma}_e \in \mathcal{B}_e^{\geq},\, e \in \mathcal{E}, \end{aligned} \tag{86}$$

where $\mathcal{B}_e^{\geq}$ and $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}$ are defined in (5) *and* (29)*, respectively.*

2) *It holds that*

$$\lim_{n\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\lambda}^{(n)}) = \hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma}^{(1)}) = F^{*}_{\mathrm{B,d,N_1}} = F^{\mathrm{alt},*}_{\mathrm{B,d,N_1}} = F^{*}_{\mathrm{B,p,N_1}} = 0, \qquad m \in \mathbb{Z}_{>0}. \tag{87}$$

Recall that $\boldsymbol{\gamma}^{(1)}$ and the sequence $\{\boldsymbol{\lambda}^{(n)}\}_n$ are specified in Definitions 33.

Proof. See Appendix I. ■

Now we focus on the optimal value $F^{*}_{\mathrm{B,p,N_1}}$ of the optimization problem in (13). We want to show that the beliefs associated with the sequence as defined in (57) converge to the location of the optimal value for the primal formulation. The first step is to show that the collection of beliefs converges to the LMP $\mathcal{B}(\mathsf{N}_1)$ as $n \to \infty$.

Lemma 36. *Consider the S-NFG* N_1 *specified in Definition 22 and Example 32. The collection of the beliefs* $\boldsymbol{b}_{f_1}^{(1,n)}$ *and* $\boldsymbol{b}_{f_2}^{(1,n)}$ *defined in* (55) *and* (56) *satisfies*

$$\lim_{n\to\infty} \boldsymbol{b}_{f_i}^{(1,n)} = \begin{pmatrix} \gamma_1^{(1)}(1) & 0 \\ 0 & \gamma_1^{(1)}(2) \end{pmatrix}, \qquad f_i \in \{f_1, f_2\},$$
$$\lim_{n\to\infty} \boldsymbol{b}_{\mathcal{F}}^{(1,n)} \in \mathcal{B}(\mathsf{N}_1). \tag{88}$$

Proof. See Appendix J. ■

By setting

$$\boldsymbol{\beta}_{f_1} = \boldsymbol{b}_{f_1}^{(1,n)}, \qquad \boldsymbol{\beta}_{f_2} = \boldsymbol{b}_{f_2}^{(1,n)},$$

where $\boldsymbol{b}_{f_1}^{(1,n)}$ and $\boldsymbol{b}_{f_2}^{(1,n)}$ are defined in (55) and (56), the objective function $F^{(1)}_{\mathrm{B,p,N_1}}$ for the primal formulation becomes

$$\begin{aligned} F^{(1)}_{\mathrm{B,p,N_1}}(\boldsymbol{b}_{\mathcal{F}}^{(1,n)}) = & -\log Z_{f_1}\big(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\lambda}^{(n)}\big) - \log Z_{f_2}\big(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\lambda}^{(n)}\big) + \sum_{x_1,x_2} b_{f_1}^{(1,n)}(x_1,x_2) \cdot (\lambda_1^{(n)}(x_1) + \lambda_2^{(n)}(x_2)) \\ & - \sum_{x_1,x_2} b_{f_2}^{(1,n)}(x_1,x_2) \cdot (\lambda_1^{(n)}(x_1) + \lambda_2^{(n)}(x_2)) \\ & + \sum_{e} \sum_{x_e :\, \gamma_1^{(1)}(x_e) > 0} \frac{b_{f_1,e}^{(1,n)}(x_e) + b_{f_2,e}^{(1,n)}(x_e)}{2} \cdot \log\!\left(\frac{2\gamma_e^{(1)}(x_e)}{b_{f_1,e}^{(1,n)}(x_e) + b_{f_2,e}^{(1,n)}(x_e)} \right), \end{aligned} \tag{89}$$

where $F^{(1)}_{\mathrm{B,p,N_1}}$ is defined in (11).

Now we want to prove that the collection of beliefs $\boldsymbol{b}_{\mathcal{F}}^{(1,n)}$ associated with the sequences $\{\boldsymbol{\lambda}^{(n)}\}_{n\in\mathbb{Z}_{>0}}$ converges to the location of the optimal value for the primal formulation, i.e, the sequence $\{\boldsymbol{b}_{\mathcal{F}}^{(1,n)}\}_{n\in\mathbb{Z}_{>0}}$ converges and $\lim_{n\to\infty} F^{(1)}_{\mathrm{B,p,N_1}}(\boldsymbol{b}_{\mathcal{F}}^{(1,n)}) = F^{\mathrm{alt},*}_{\mathrm{B,d,N_1}} = 0$. By (87) and the expression of $F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}$ in (54), we know that the first two terms in (89) converge to zero as $n \to \infty$. It is sufficient to prove that the remaining terms in (89) converge to zero as well.

Lemma 37. *Considering the sequence* $\boldsymbol{\lambda}^{(n)}$ *given in Definition 33, we obtain*

$$\lim_{n\to\infty} \left(\sum_{x_1,x_2} b_{f_1}^{(1,n)}(x_1,x_2) \cdot \Big(\lambda_1^{(n)}(x_1) + \lambda_2^{(n)}(x_2)\Big) - \sum_{x_1,x_2} b_{f_2}^{(1,n)}(x_1,x_2) \cdot \Big(\lambda_1^{(n)}(x_1) + \lambda_2^{(n)}(x_2)\Big) \right) = 0.$$

Proof. See Appendix K. ■

Theorem 38. *For the sequence* $\boldsymbol{\lambda}^{(n)}$ *specified in Definition 33, it holds that*

$$\lim_{n\to\infty} F^{(1)}_{\mathrm{B,p,N_1}}(\boldsymbol{b}_{\mathcal{F}}^{(1,n)}) = F^{\mathrm{alt},*}_{\mathrm{B,d,N_1}} = F^{*}_{\mathrm{B,d,N_1}} = F^{*}_{\mathrm{B,p,N_1}} = 0, \tag{90}$$

where $\boldsymbol{b}_{f_1}^{(1,n)}$ *and* $\boldsymbol{b}_{f_2}^{(1,n)}$ *are given in* (55) *and* (56)*, the function* $F^{(1)}_{\mathrm{B,p,N_1}}$ *is given in* (11)*, the quantity* $F^{\mathrm{alt},*}_{\mathrm{B,d,N_1}}$ *is defined in* (22)*, the quantity* $F^{*}_{\mathrm{B,d,N_1}}$ *is defined in* (20)*, and the quantity* $F^{*}_{\mathrm{B,p,N_1}}$ *is the minimum of the constrained Bethe partition function as defined in* (13).

Proof. See Appendix L. ■

After relating the sequences $\boldsymbol{\gamma}^{(1)}$ and $\{\boldsymbol{\lambda}^{(n)}\}_n$ to the locations of the optimal value for the primal and dual formulations, we want to relate these sequences to an SPA fixed point for the considered S-NFG. We first define messages based on the sequences.

Definition 39. *For each $e = (f_1, f_2) \in [2]$, we define*

$$\mu_{e\to f_i}^{(1,n)}(x_e) \triangleq \exp\bigl(\lambda_{e,f_i}^{(n)}(x_e)\bigr) \cdot \sqrt{\gamma_e^{(1)}(x_e)}, \quad x_e \in \mathcal{X}_e,\, i \in [2]. \tag{91}$$

$$Z_{\mu_{e\to f_i}}^{(1,n)} \triangleq \sum_{x_e \in \mathcal{X}_e} \mu_{e\to f_i}^{(1,n)}(x_e), \qquad i \in [2], \tag{92}$$

$$\mu_{e\to f_1,\mathrm{SPA}}^{(1,n)}(x_e) \triangleq \frac{1}{C_{e\to f_1}^{(1,n)}} \cdot \frac{b_{f_2,e}^{(1,n)}(x_e)}{\mu_{e\to f_2}^{(1,n)}(x_e)}, \qquad \mu_{e\to f_2,\mathrm{SPA}}^{(1,n)}(x_e) \triangleq \frac{1}{C_{e\to f_2}^{(1,n)}} \cdot \frac{b_{f_1,e}^{(1,n)}(x_e)}{\mu_{e\to f_1}^{(1,n)}(x_e)}, \tag{93}$$

where $C_{e\to f_1}^{(1,n)}$ and $C_{e\to f_2}^{(1,n)}$ are normalization constants:

$$C_{e\to f_1}^{(1,n)} \triangleq \sum_{x_e} \frac{b_{f_2,e}^{(1,n)}(x_e)}{\mu_{e\to f_2}^{(1,n)}(x_e)}, \qquad C_{e\to f_2}^{(1,n)} \triangleq \sum_{x_e} \frac{b_{f_1,e}^{(1,n)}(x_e)}{\mu_{e\to f_1}^{(1,n)}(x_e)}, \tag{94}$$

and the marginal $b_{f,e}$ for $f \in \{f_1, f_2\}$ and $e \in [2]$ is defined in (64)*.*

Theorem 40. *The following equation for SPA fixed-point messages holds:*

$$\lim_{n\to\infty} \mu_{e\to f_i,\mathrm{SPA}}^{(1,n)}(x_e) = \lim_{n\to\infty} \frac{\mu_{e\to f_i}^{(1,n)}(x_e)}{Z_{\mu_{e\to f_i}}^{(1,n)}} \in \mathbb{R}_{\geq 0}, \qquad x_e \in \mathcal{X}_e,\, e = (f_1, f_2) \in [2],\, i \in [2], \tag{95}$$

where the vector $\mu_{e\to f_i}^{(1,n)}$ is defined in (91)*, the constant $Z_{\mu_{e\to f_i}}^{(1,n)}$ is defined in* (92)*, and messages $\mu_{e\to f_1,\mathrm{SPA}}^{(1,n)}$ and $\mu_{e\to f_2,\mathrm{SPA}}^{(m,n)}$ are defined in* (93)*. It shows that the message sequence $\boldsymbol{\mu}^{(1,n)} \triangleq \{\boldsymbol{\mu}_{e\to f}^{(1,n)}\}_{e\in\partial f, f\in\mathcal{F}}$ converges to an SPA fixed point as $n \to \infty$. In particular, it holds that*

$$\lim_{n\to\infty}\Bigl(\mu_{1\to f_1,\mathrm{SPA}}^{(1,n)}(1) \quad \mu_{1\to f_1,\mathrm{SPA}}^{(1,n)}(2)\Bigr)^{\mathsf{T}} = \Bigl(0, \quad 1\Bigr)^{\mathsf{T}}, \qquad \lim_{n\to\infty}\Bigl(\mu_{2\to f_1,\mathrm{SPA}}^{(1,n)}(1) \quad \mu_{2\to f_1,\mathrm{SPA}}^{(1,n)}(2)\Bigr)^{\mathsf{T}} = \Bigl(1, \quad 0\Bigr)^{\mathsf{T}},$$

which, by Example 32, shows that the vectors $(\mu_{1\to f_1,\mathrm{SPA}}^{(1,n)}(x_e))_{x_e\in\mathcal{X}_e}$ and $(\mu_{2\to f_1,\mathrm{SPA}}^{(1,n)}(x_e))_{x_e\in\mathcal{X}_e}$ converge to the left and right eigenvectors of the matrix $\boldsymbol{f}_1 \cdot \boldsymbol{f}_2^{\mathsf{T}}$, respectively, where $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ are given in Example 32.

Proof. For the sequence $\boldsymbol{\lambda}^{(n)}$ specified in Definition 33, it holds that

$$\lim_{n\to\infty} \frac{\mu_{1\to f_1}^{(1,n)}(1)}{Z_{\mu_{1\to f_1}}^{(1,n)}} = \lim_{n\to\infty} \frac{\exp(-n) \cdot \sqrt{\gamma_1^{(1)}(1)}}{\exp(-n) \cdot \sqrt{\gamma_1^{(1)}(1)} + \sqrt{\gamma_1^{(1)}(2)}} = 0.$$

Similarly, we obtain

$$\lim_{n\to\infty} \frac{\mu_{1\to f_1}^{(1,n)}(2)}{Z_{\mu_{1\to f_1}}^{(1,n)}} = 1, \qquad \lim_{n\to\infty} \frac{\mu_{2\to f_1}^{(1,n)}(1)}{Z_{\mu_{2\to f_1}}^{(1,n)}} = 1, \qquad \lim_{n\to\infty} \frac{\mu_{2\to f_1}^{(1,n)}(2)}{Z_{\mu_{2\to f_1}}^{(1,n)}} = 0,$$
$$\lim_{n\to\infty} \frac{\mu_{1\to f_2}^{(1,n)}(1)}{Z_{\mu_{1\to f_2}}^{(1,n)}} = 1, \qquad \lim_{n\to\infty} \frac{\mu_{1\to f_2}^{(1,n)}(2)}{Z_{\mu_{1\to f_2}}^{(1,n)}} = 0, \qquad \lim_{n\to\infty} \frac{\mu_{2\to f_2}^{(1,n)}(1)}{Z_{\mu_{2\to f_2}}^{(1,n)}} = 0, \qquad \lim_{n\to\infty} \frac{\mu_{2\to f_2}^{(1,n)}(2)}{Z_{\mu_{2\to f_2}}^{(1,n)}} = 1.$$

Also by the definitions of $b_{f_1}^{(1,n)}$ and $\mu_{e\to f,\mathrm{SPA}}^{(1,n)}$ in (55) and (93), we have

$$\lim_{n\to\infty}\Bigl(\mu_{1\to f_1,\mathrm{SPA}}^{(1,n)}(1) \quad \mu_{1\to f_1,\mathrm{SPA}}^{(1,n)}(2)\Bigr) = \lim_{n\to\infty}\Bigl(\mu_{1\to f_2,\mathrm{SPA}}^{(1,n)}(1) \quad \mu_{1\to f_2,\mathrm{SPA}}^{(1,n)}(2)\Bigr) = \Bigl(0, \quad 1\Bigr),$$
$$\lim_{n\to\infty}\Bigl(\mu_{2\to f_1,\mathrm{SPA}}^{(1,n)}(1) \quad \mu_{2\to f_1,\mathrm{SPA}}^{(1,n)}(2)\Bigr) = \lim_{n\to\infty}\Bigl(\mu_{2\to f_2,\mathrm{SPA}}^{(1,n)}(1) \quad \mu_{2\to f_2,\mathrm{SPA}}^{(1,n)}(2)\Bigr) = \Bigl(1, \quad 0\Bigr).$$

The derivations of the above expressions are similar and thus they are omitted here. ■

Example 41. *We present some numerical results w.r.t. $F_{\mathrm{B,d,N_1}}^{\mathrm{alt}}(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ and $F_{\mathrm{B,p,N_1}}(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ for corroborating the theoretical results obtained in this section.*

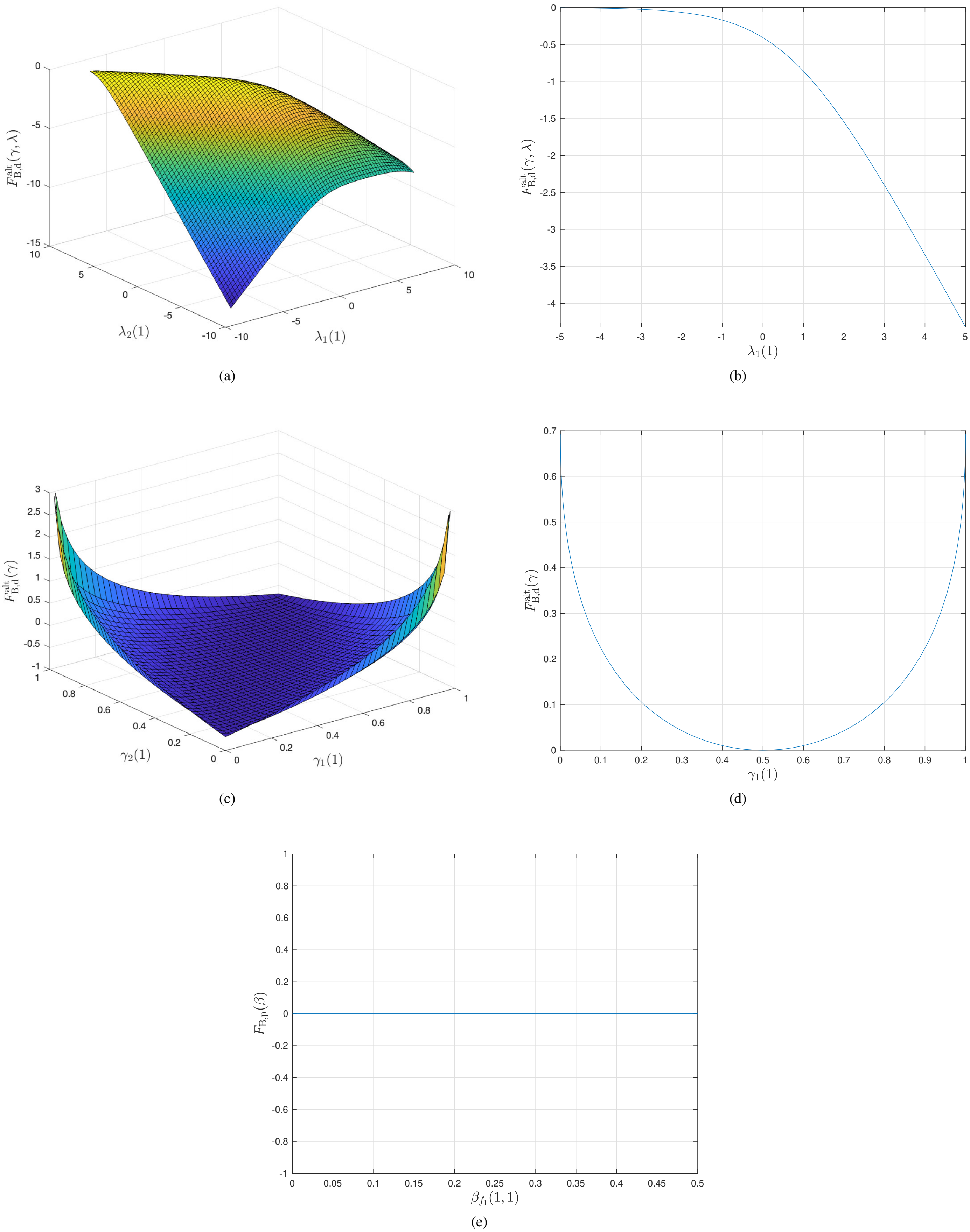


Fig. 4: The numerical results in Example 41.

1) *The shape of* $F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma},\boldsymbol{\lambda})$ *given in* (54) *for* $\gamma_e(x_e)=1/2$ *for all* $x_e\in\mathcal{X}_e$ *and* $\lambda_e(2)=0$ *for all* $e\in[2]$, *is plotted in Fig. 4(a). In this case, the function* $F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma},\boldsymbol{\lambda})$ *equals*

$$F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma},\boldsymbol{\lambda})=2\log 2-\log\Bigl(1+\exp(\lambda_1(1)+\lambda_2(1))+\exp(\lambda_1(1))\Bigr)-\log\Bigl(1+\exp(\lambda_1(1)+\lambda_2(1))\Bigr).$$

We can see that the function $F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma},\boldsymbol{\lambda})$ *is concave w.r.t.* $\boldsymbol{\lambda}$, *which corroborates Item 1 in Theorem 14.*

2) *Fig. 4(b) plots* $F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma},\boldsymbol{\lambda})$ *w.r.t.* $\lambda_1(1)$ *under the same condition as the condition in Fig. 4(a) with an additional condition* $\lambda_2(1)=-\lambda_1(1)$. *In this case, the function* $F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma},\boldsymbol{\lambda})$ *equals*

$$F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma},\boldsymbol{\lambda})=\log 2-\log\Bigl(2+\exp(\lambda_1(1))\Bigr).$$

From the figure, we can see that the function is decreasing w.r.t. $\lambda_1(1)$, *which corroborates the expression in* (86) *that* $F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma},\boldsymbol{\lambda})$ *takes supremum when* $\lambda_1(1)\to-\infty$.

3) *The shape of* $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma})$ *given in* (86), *is plotted in Fig. 4(c). By* (86), *the function* $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma})$ *equals*

$$\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma})=-2\log\Bigl(\sqrt{\gamma_1(1)\cdot\gamma_2(1)}+\sqrt{\bigl(1-\gamma_1(1)\bigr)\cdot\bigl(1-\gamma_2(1)\bigr)}\Bigr).$$

We can see that the function $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma})$ *is convex w.r.t.* $\boldsymbol{\gamma}$, *which corroborates Item 2 in Theorem 14.*

4) *In particular, Fig. 4(d) plots* $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma})$ *w.r.t.* $\gamma_1(1)$ *when* $\gamma_2(1)=1/2$. *In this case, the function* $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma})$ *equals*

$$\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma})=-2\log\Bigl(\sqrt{\frac{1}{2}\gamma_1(1)}+\sqrt{\frac{1}{2}\bigl(1-\gamma_1(1)\bigr)}\Bigr).$$

From the figure, we can see that the function takes minimum when $\gamma_1(1)=\gamma_2(1)=1/2$, *which corroborates the equalities in* (87) *that* $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma})$ *takes minimum when* $\gamma_1(1)=\gamma_2(1)$.

5) *The shape of the Bethe free energy function* $F_{\mathrm{B,p,N_1}}(\boldsymbol{\beta})$ *given in* (81) *is plotted in Fig. 4(e), which corroborates the expression of* $F_{\mathrm{B,p,N_1}}$ *in* (81).

We also discuss the behavior of the double-loop algorithm applied in the considered S-NFG N_1.

Proposition 42. *Consider* N_1 *specified in Definition 22 and Example 32. If we run the double-loop algorithm defined in Algorithm 1 for* N_1 *and set*

$$\boldsymbol{\gamma}^{(0)}_{\mathrm{dl}}\in\prod_{e\in\mathcal{E}}\mathcal{B}^{>}_{e},$$

then we have

$$\begin{aligned}
\lim_{t_1\to\infty}\Bigl(\exp\Bigl(\lambda^{(t_1)}_{\mathrm{dl},1,f_1}(1)\Bigr),\quad \exp\Bigl(\lambda^{(t_1)}_{\mathrm{dl},1,f_1}(2)\Bigr)\Bigr)&=\lim_{t_1\to\infty}\Bigl(\exp\Bigl(\lambda^{(t_1)}_{\mathrm{dl},2,f_2}(1)\Bigr),\quad \exp\Bigl(\lambda^{(t_1)}_{\mathrm{dl},2,f_2}(2)\Bigr)\Bigr)=\Bigl(0,\quad 1\Bigr),\\
\lim_{t_1\to\infty}\Bigl(\exp\Bigl(\lambda^{(t_1)}_{\mathrm{dl},1,f_2}(1)\Bigr),\quad \exp\Bigl(\lambda^{(t_1)}_{\mathrm{dl},1,f_2}(2)\Bigr)\Bigr)&=\lim_{t_1\to\infty}\Bigl(\exp\Bigl(\lambda^{(t_1)}_{\mathrm{dl},2,f_1}(1)\Bigr),\quad \exp\Bigl(\lambda^{(t_1)}_{\mathrm{dl},2,f_1}(2)\Bigr)\Bigr)=\Bigl(1,\quad 0\Bigr),\\
\boldsymbol{\gamma}^{(t_2)}_{\mathrm{dl}}=\boldsymbol{\gamma}^{(0)}_{\mathrm{dl}},&\qquad t_2\in\mathbb{Z}_{>0}.
\end{aligned}$$

Proof. It can be proven straightforwardly. ∎

V. The Analysis of a Double-Cycle S-NFG Example

In this section, we consider an example double-cycle S-NFG N_2 as shown in Fig. 5.

Remark 43. *We make the following remarks for* N_2 *as shown in Fig. 5.*

1) *The alphabet of the variables is given by* $\mathcal{X}_e=[2]$ *for all* $e\in\mathcal{E}$.

2) *The local functions satisfy*

$$f_1(x_1, x_2, x_3) = f_2(x_1, x_2, x_3) = [x_1 = x_2 = x_3], \qquad x_1, x_2, x_3 \in [2].$$

3) *The associated Bethe free energy function* $F_{\mathrm{B,p,N_2}}$ *defined in* (10) *equals*

$$F_{\mathrm{B,p,N_2}}(\boldsymbol{\beta}) = -H_{\mathrm{B},f}(\boldsymbol{\beta}_{f_1}) - H_{\mathrm{B},f}(\boldsymbol{\beta}_{f_2}) + \sum_e H_{\mathrm{B},e}(\boldsymbol{\beta}_e) \overset{(a)}{=} -\sum_{x_1} \beta_e(x_1) \cdot \log \beta_e(x_1), \qquad \boldsymbol{\beta} \in \mathcal{B}(\mathsf{N}_2),\ e \in [3].$$

where $\mathcal{B}(\mathsf{N}_2)$ *is defined in* (7)*, and where step* (a) *follows from*

$$\begin{aligned} H_{\mathrm{B},f}(\boldsymbol{\beta}_f) &= -\sum_{\boldsymbol{x}_f \in \mathcal{X}_f} \beta_f(\boldsymbol{x}_f) \cdot \log \beta_f(\boldsymbol{x}_f) \\ &= -\sum_{\boldsymbol{x}_f :\, f(\boldsymbol{x}_f) > 0} \beta_f(\boldsymbol{x}_f) \cdot \log \beta_f(\boldsymbol{x}_f) \\ &= -\sum_{x_e} \beta_f\big(x_e, x_e, x_e\big) \cdot \log \beta_f\big(x_e, x_e, x_e\big) \\ &= -\sum_{x_e} \beta_e(x_e) \cdot \log \beta_e(x_e), \qquad \boldsymbol{\beta} \in \mathcal{B}(\mathsf{N}_2). \end{aligned}$$

Note that $F_{\mathrm{B,p,N_2}}(\boldsymbol{\beta})$ *is a concave, not a convex, function of* $\boldsymbol{\beta}$*. The above expression implies*

$$\min_{\boldsymbol{\beta} \in \mathcal{B}(\mathsf{N}_2)} F_{\mathrm{B,p,N_2}}(\boldsymbol{\beta}) = 0.$$

4) *The S-NFG* N_2 *has three SPA fixed points:*
   a) $\boldsymbol{\mu}_{e \to f_1} = \boldsymbol{\mu}_{e \to f_2} = \begin{pmatrix} 1, & 0 \end{pmatrix}^{\mathsf{T}}$*, for all* $e \in \mathcal{E}$*;*
   b) $\boldsymbol{\mu}_{e \to f_1} = \boldsymbol{\mu}_{e \to f_2} = \begin{pmatrix} 0, & 1 \end{pmatrix}^{\mathsf{T}}$*, for all* $e \in \mathcal{E}$*;*
   c) $\boldsymbol{\mu}_{e \to f_1} = \boldsymbol{\mu}_{e \to f_2} = \frac{1}{2}\begin{pmatrix} 1, & 1 \end{pmatrix}^{\mathsf{T}}$*, for all* $e \in \mathcal{E}$*.*

   The first two fixed points correspond to the minima of $F_{\mathrm{B,p,N_2}}(\boldsymbol{\beta})$*,* $\boldsymbol{\beta} \in \mathcal{B}(\mathsf{N}_1)$*. The last fixed point corresponds to the maximum of* $F_{\mathrm{B,p,N_2}}(\boldsymbol{\beta})$*,* $\boldsymbol{\beta} \in \mathcal{B}(\mathsf{N}_1)$*.*

5) *The function* Z_{f_1} *defined in* (16) *can be written as*

$$\begin{aligned} Z_{f_1}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) &= \sum_{\boldsymbol{x}} [x_1 = x_2 = x_3] \cdot \exp\Big(\sum_e \lambda_{e,f_1}(x_e)\Big) \cdot \sqrt{\prod_e \gamma_e(x_e)} \\ &\overset{(a)}{=} \sum_{x_1} \exp\Big(\sum_e \lambda_e(x_1)\Big) \cdot \sqrt{\prod_e \gamma_e(x_1)}, \end{aligned}$$

where step (a) *follows from the definition of* $\boldsymbol{\lambda}_{e,f}$ *in* (15)*. Similarly, the function* Z_{f_2} *can be written as*

$$Z_{f_2}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) = \sum_{x_1} \exp\Big(-\sum_e \lambda_e(x_1)\Big) \cdot \sqrt{\prod_e \gamma_e(x_1)}. \tag{96}$$

6) *The vector* $\boldsymbol{\lambda} \in \operatorname{argmax}_{\boldsymbol{\lambda} \in \mathbb{R}^{|\mathcal{X}|}} F^{\mathrm{alt}}_{\mathrm{B,d,N_2}}(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ *satisfies*

$$\sum_e \lambda_e(x_1) = 0, \qquad x_1 \in [2],$$

which can be proven directly from Theorem 14, i.e., $F^{\mathrm{alt}}_{\mathrm{B,d,N_2}}(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ *is convex w.r.t.* $\boldsymbol{\lambda}$ *and the necessary optimality conditions in [19, Proposition 1.1.1]: if* $\boldsymbol{\lambda}$ *is at the location of the optimal value, then* $\boldsymbol{\lambda}$ *is at a stationary point of* $F^{\mathrm{alt}}_{\mathrm{B,d,N_2}}(\boldsymbol{\gamma}, \boldsymbol{\lambda})$*.*

7) *By the previous property, we have*

$$\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_2}}(\boldsymbol{\gamma}) = -2 \log\left(\sum_{x_1} \sqrt{\prod_e \gamma_e(x_1)}\right).$$

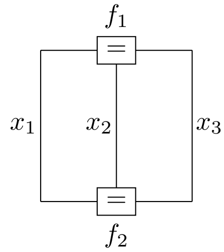


Fig. 5: S-NFG N_2.

In particular, it holds that

$$\begin{aligned}\sum_{x_1}\sqrt{\prod_e \gamma_e(x_1)} &= \sqrt{\gamma_1(1)\cdot\gamma_2(1)\cdot\gamma_3(1)} + \sqrt{\gamma_1(2)\cdot\gamma_2(2)\cdot\gamma_3(2)} \\ &\overset{(a)}{\leq} \sqrt{\gamma_1(1)+\gamma_1(2)}\cdot\Big(\sqrt{\gamma_2(1)\cdot\gamma_3(1)}+\sqrt{\gamma_2(2)\cdot\gamma_3(2)}\Big) \\ &\overset{(b)}{=} \sqrt{\gamma_2(1)\cdot\gamma_3(1)}+\sqrt{\gamma_2(2)\cdot\gamma_3(2)} \\ &\overset{(c)}{\leq} \sqrt{\gamma_2(1)+\gamma_2(2)}\cdot\sqrt{\gamma_3(1)+\gamma_3(2)} \\ &\overset{(d)}{=} 1, \qquad \boldsymbol{\gamma}_e \in \mathcal{B}_e^{\geq},\, e\in\mathcal{E},\end{aligned}$$

where step (a) follows from the fact that $\gamma_1(1)$ and $\gamma_1(2)$ are non-negative. where step (b) follows from $\boldsymbol{\gamma}_1 \in \mathcal{B}_1^{\geq}$, i.e., $\gamma_1(1)+\gamma_1(2)=1$, where step (c) follows from the Cauchy-Schwarz inequality, and where step (d) follows from $\boldsymbol{\gamma}_2 \in \mathcal{B}_2^{\geq}$ and $\boldsymbol{\gamma}_3 \in \mathcal{B}_3^{\geq}$ i.e., $\gamma_2(1)+\gamma_2(2)=\gamma_3(1)+\gamma_3(2)=1$. It further implies

$$\hat{F}^{\mathrm{alt}}_{\mathrm{B,d},\mathsf{N}_2}(\boldsymbol{\gamma}) \geq 0.$$

When

$$\big(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_3\big) = \Big(\big(1,\ \ 0\big)^{\mathsf{T}},\ \big(1,\ \ 0\big)^{\mathsf{T}},\ \big(1,\ \ 0\big)^{\mathsf{T}}\Big)\ \textit{or}\ \Big(\big(0,\ \ 1\big)^{\mathsf{T}},\ \big(0,\ \ 1\big)^{\mathsf{T}},\ \big(0,\ \ 1\big)^{\mathsf{T}}\Big),$$

we have $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d},\mathsf{N}_2}(\boldsymbol{\gamma})=0$, which implies

$$\min_{\boldsymbol{\gamma}} \hat{F}^{\mathrm{alt}}_{\mathrm{B,d},\mathsf{N}_2}(\boldsymbol{\gamma}) = 0, \qquad \text{s.t.}\quad \boldsymbol{\gamma}\in\prod_e \mathcal{B}_e^{\geq}.$$

8) *For the S-NFG N_2 considered in this section, the possible sequences $\{\boldsymbol{\gamma}^{(m)}\}_m$ defined in Item 2 in Definition 17 are*

$$\big(\boldsymbol{\gamma}_1^{(m)},\boldsymbol{\gamma}_2^{(m)},\boldsymbol{\gamma}_3^{(m)}\big) = \Big(\big(1,\ \ 0\big)^{\mathsf{T}},\ \big(1,\ \ 0\big)^{\mathsf{T}},\ \big(1,\ \ 0\big)^{\mathsf{T}}\Big)\ \textit{or}\ \Big(\big(0,\ \ 1\big)^{\mathsf{T}},\ \big(0,\ \ 1\big)^{\mathsf{T}},\ \big(0,\ \ 1\big)^{\mathsf{T}}\Big), \qquad m\in\mathbb{Z}_{>0}.$$

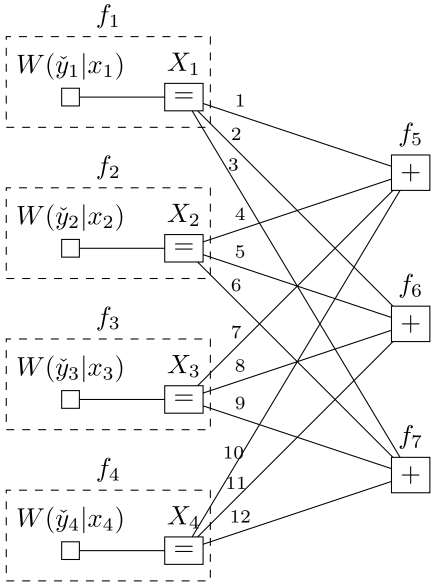
9) *For the S-NFG N_2 considered in this section, the possible sequences $\{\boldsymbol{\lambda}^{(n)}\}_n$ defined in Item 3 in Definition 17 are arbitrary sequences satisfying*

$$\sum_e \lambda_e^{(n)}(x_1) = 0, \qquad x_1\in[2],\, n\in\mathbb{Z}_{>0}.$$

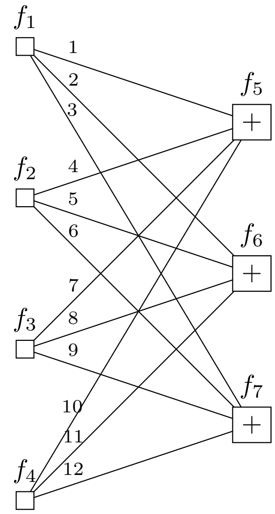
VI. The Analysis of a Low-Density Parity-Check (LDPC) Code Example

In this section, we consider the example S-NFG N_3 representing a $(3,4)$-regular LDPC code based on the parity-check matrix

$$H = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$



(a) S-NFG N_3.



(b) S-NFG N_4.

Fig. 6: The S-NFGs in Definition 44.

as shown in Fig. 6(a), that is used for data transmission over a memoryless channel with input alphabet $x \in \{0,1\}$, output alphabet $y \in \{0,1\}$, and channel law $W(y|x)$ with $W(y|x) \in \mathbb{R}_{\geq 0}$ for all x, y and $\sum_y W(y|x) = 1$ for all x.

The S-NFG N_4 in Fig. 6(b) is obtained from N_3 by applying the closing-the-box (CTB) operation in N_3. Consider the dashed boxes in Fig. 6(a). Its exterior function is defined to be the sum, over the internal variables, of the product of the internal local functions. Replacing this dashed box by a single function node that represents this exterior function is known as the CTB operation [3]. For details of this operation in N_3, see the definition in (98).

Definition 44. *The details of N_3 and N_4 as shown in Figures 6(a) and 6(b) are given as follows.*

1) *The set of the edges in the S-NFG N_4 is $\mathcal{E} = [12]$.*
2) *The alphabet of the variables x_e is set to be $\mathcal{X}_e = \{0,1\}$ for $e \in \mathcal{E}$.*
3) *The variables $y_1, \ldots, y_4$ take values in $\{0,1\}$.*
4) *The observed variables $\check{y}_1, \ldots, \check{y}_4$ take values in $\{0,1\}$.*
5) *The set of the function nodes on the LHS in N_4 is given by $\{f_i\}_{i \in \mathcal{I}}$ with $\mathcal{I} = [4]$.*
6) *The channel law is defined to be an arbitrary function such that*

$$W(y|x) \in \mathbb{R}_{>0},\, x, y \in \{0,1\}, \qquad \sum_y W(y|x) = 1,\, x \in \{0,1\}.$$

7) *For each $i \in \mathcal{I}$, the conditional probability $P_{Y_i|X_i}$ is defined to be*

$$P_{Y_i|X_i}(y_i|x_i) \triangleq W(y_i|x_i), \qquad x_i, y_i \in \{0,1\}.$$

8) *The set of the function nodes on the RHS in N_4 is given by $\{f_j\}_{j \in \mathcal{J}}$ with $\mathcal{J} = \{5,6,7\}$.*
9) *For each $j \in \mathcal{J}$, the alphabet for the parity-check node f_j is given by*

$$\mathcal{X}_{f_j} \triangleq \left\{ (x_1, \ldots, x_4) \in \{0,1\}^4 \;\middle|\; x_1 + \cdots + x_4 = 0 \mod 2 \right\}. \tag{97}$$

10) *For each $i \in \mathcal{I}$ and fixed y_e, $e \in \partial f$, the function f_i is defined to be*

$$f_i(x_{\partial f_i}) \triangleq [\textit{all } x_e,\, e \in \partial f_i \textit{ are equal}] \cdot \prod_{e \in \partial f_i} W(y_e|x_e), \qquad x_e \in \{0,1\}. \tag{98}$$

Note that $f_i(x_i)$ can be obtained by applying the CTB operation in Fig. 6(a) for each $i \in \mathcal{I}$.

In the remaining part of this section, if there is no ambiguity, we use the shorthands $\sum_i$, $\prod_i$, $\sum_{\boldsymbol{x}_{f_i}}$, $\sum_j$, $\prod_j$ and $\sum_{\boldsymbol{x}_{f_j}}$ for $\sum_{i\in\mathcal{I}}$, $\prod_{i\in\mathcal{I}}$, $\sum_{\boldsymbol{x}_{f_i}\in\{0,1\}}$, $\sum_{j\in\mathcal{J}}$, $\prod_{j\in\mathcal{J}}$, and $\sum_{\boldsymbol{x}_{f_j}\in\mathcal{X}_{f_j}}$, respectively.

Assumption 45. *We assume the observed variables are fixed to be $\check{y}_1 = \cdots = \check{y}_4 = 0$ and the channel law satisfies*

$$W(0|x) = W(1|1-x), \qquad x \in \{0,1\}, \qquad W(0|0) > W(0|1) > 0, \tag{99}$$

$$\big(W(0|0)\big)^{4/3} \geq \sum_{\boldsymbol{x}_{f_i}\in\mathcal{X}_{f_i}\setminus\{(0,\ldots,0)\}} \prod_{e\in\partial f_i} \big(W(0|x_e)\big)^{1/3} = 6\big(W(0|0)\cdot W(0|1)\big)^{2/3} + \big(W(0|1)\big)^{4/3}, \qquad i \in \mathcal{I}. \tag{100}$$

In the following, we will see that if the channel law satisfies Assumption 45, then the location of the optimal value for the associated primal formulation has a simple structure. In order to avoiding confusion, we will discuss only S-NFG N_4 in the remaining part of this section. Note that the Bethe free energy functions for N_3 and N_4 are defined over different LMPs $\mathcal{B}(\mathsf{N}_3)$ and $\mathcal{B}(\mathsf{N}_4)$, respectively.

The Bethe free energy function $F_{\mathrm{B,p},\mathsf{N}_4}(\boldsymbol{\beta})$, as defined in in (10), is given by

$$F_{\mathrm{B,p},\mathsf{N}_4}(\boldsymbol{\beta}) = \sum_j \sum_{\boldsymbol{x}_{f_j}} \beta_{f_j}(\boldsymbol{x}_{f_j}) \cdot \log \beta_{f_j}(\boldsymbol{x}_{f_j}) - \sum_i \sum_{\boldsymbol{x}_{f_i}} \beta_{f_i}(\boldsymbol{x}_{f_i}) \cdot \log f_i(\boldsymbol{x}_{f_i}) - 2\sum_i \sum_{\boldsymbol{x}_{f_i}} \beta_{f_i}(\boldsymbol{x}_{f_i}) \cdot \log \beta_{f_i}(\boldsymbol{x}_{f_i}). \tag{101}$$

The associated Bethe partition function $Z^*_{\mathrm{B,p},\mathsf{N}_4}$ defined in (12) is given by

$$Z^*_{\mathrm{B,p},\mathsf{N}_4} = \exp\Big(- \min_{\boldsymbol{\beta}\in\mathcal{B}(\mathsf{N}_4)} F_{\mathrm{B,p},\mathsf{N}_4}(\boldsymbol{\beta}) \Big). \tag{102}$$

where the LMP $\mathcal{B}(\mathsf{N}_4)$ is defined in (7).

We want to show that

$$Z^*_{\mathrm{B,p},\mathsf{N}_4} \overset{(a)}{=} Z^*_{\mathrm{B,d},\mathsf{N}_4} \leq \prod_i W(0|0), \tag{103}$$

where step (a) follows from Proposition 16. Recall that $Z^*_{\mathrm{B,d},\mathsf{N}_4}$ defined in (19) equals

$$\begin{aligned} Z^*_{\mathrm{B,d},\mathsf{N}_4} &= \sup_{\boldsymbol{\gamma}} \inf_{\boldsymbol{\lambda}} Z^*_{\mathrm{B,d},\mathsf{N}_4}(\boldsymbol{\gamma},\boldsymbol{\lambda}) \\ &\text{s.t.} \quad \boldsymbol{\gamma}_e \in \mathcal{B}_e^{>},\ \boldsymbol{\lambda}_e \in \mathbb{R}^{|\mathcal{X}_e|}, \qquad e \in \mathcal{E}, \end{aligned} \tag{104}$$

where $\mathcal{B}_e^{>}$ is defined in (6) for each $e \in \mathcal{E}$, and where the functions $Z_{\mathrm{B,d},\mathsf{N}_4}(\boldsymbol{\gamma},\boldsymbol{\lambda})$ and Z_f, as defined in (18) and (16), respectively, satisfies

$$Z_{\mathrm{B,d},\mathsf{N}_4}(\boldsymbol{\gamma},\boldsymbol{\lambda}) = \bigg(\prod_i Z_{f_i}(\boldsymbol{\gamma}_{\partial f_i}, \boldsymbol{\lambda}_{\partial f i}) \bigg) \cdot \bigg(\prod_j Z_{f_j}(\boldsymbol{\gamma}_{\partial f_j}, \boldsymbol{\lambda}_{\partial f j}) \bigg), \tag{105}$$

$$Z_{f_i}(\boldsymbol{\gamma}_{\partial f i}, \boldsymbol{\lambda}_{\partial f i}) = \sum_{\boldsymbol{x}_{f_i}} [\text{all } x_e,\, e\in\partial f_i \text{ are equal}] \cdot \prod_{e\in\partial f_i} \Big(\big(W(0|x_e)\big)^{1/3} \cdot \exp(\lambda_e(x_e)) \cdot \sqrt{\gamma_e(x_e)} \Big), \qquad i \in \mathcal{I}, \tag{106}$$

$$Z_{f_j}(\boldsymbol{\gamma}_{\partial f_j}, \boldsymbol{\lambda}_{\partial f_j}) = \sum_{\boldsymbol{x}_{f_j}} \Bigg[\sum_{e\in\partial f} x_e = 0 \mod 2 \Bigg] \cdot \prod_{e\in\partial f_j} \Big(\exp(-\lambda_e(x_e)) \cdot \sqrt{\gamma_e(x_e)} \Big), \qquad j \in \mathcal{J}. \tag{107}$$

For comparison, we also consider the function $Z^{\mathrm{alt},*}_{\mathrm{B,d},\mathsf{N}_4}$ defined in (21), which is given by

$$\begin{aligned} Z^{\mathrm{alt},*}_{\mathrm{B,d},\mathsf{N}_4} &= \sup_{\boldsymbol{\gamma}} \inf_{\boldsymbol{\lambda}} Z^*_{\mathrm{B,d},\mathsf{N}_4}(\boldsymbol{\gamma},\boldsymbol{\lambda}) \\ &\text{s.t.} \quad \boldsymbol{\gamma}_e \in \mathcal{B}_e^{\geq},\ \boldsymbol{\lambda}_e \in \mathbb{R}^{|\mathcal{X}_e|}, \qquad e \in \mathcal{E}, \end{aligned} \tag{108}$$

where $\mathcal{B}_e^{\geq}$ is defined in (6). By Proposition 15, to prove (103), it is sufficent to show that

$$Z^*_{\mathrm{B,p},\mathsf{N}_4} = Z^*_{\mathrm{B,d},\mathsf{N}_4} \leq Z^{\mathrm{alt},*}_{\mathrm{B,d},\mathsf{N}_4} \leq \prod_i W(0|0).$$

Lemma 46. *Consider the S-NFG* N_4 *specified in Definition 44 under Assumption 45. It holds that*

$$Z_{\mathrm{B,d,N_4}}^{\mathrm{alt},*} \leq \prod_i W(0|0),$$

where $Z_{\mathrm{B,d,N_4}}^{\mathrm{alt},*}$ *is given in* (108).

Proof. See Appendix M. ■

Theorem 47. *Consider the S-NFG* N_4 *specified in Definition 44 under Assumption 45. We have*

$$Z_{\mathrm{B,p,N_4}}^{*} = Z_{\mathrm{B,d,N_4}}^{*} = Z_{\mathrm{B,d,N_4}}^{\mathrm{alt},*} = \Big(W(0|0)\Big)^4,$$

where $Z_{\mathrm{B,p,N_4}}^{*}$ *and* $Z_{\mathrm{B,d,N_4}}^{*}$ *are given in* (102) *and* (104)*, respectively. Also the following set of beliefs corresponds to the location of the optimal value* $Z_{\mathrm{B,p,N_4}}^{*}$ *of the optimzation problem in* (102).

$$\beta_e(x_e) = \beta_{f_i}(x_e) = \begin{cases} 1 & x_e = 0 \\ 0 & x_e = 1 \end{cases}, \qquad i \in \mathcal{I},\, e \in \mathcal{E}, \qquad \beta_{f_j}(\boldsymbol{x}_{f_j}) = \begin{cases} 1 & \boldsymbol{x}_{f_j} = \boldsymbol{0} \\ 0 & \textit{Otherwise} \end{cases}, \qquad \boldsymbol{x}_{f_j} \in \mathcal{X}_{f_j},\, j \in \mathcal{J}. \tag{109}$$

Proof. By Propositions 15 and 16 and Lemma 46, we know that

$$Z_{\mathrm{B,d,N_4}}^{*} = Z_{\mathrm{B,p,N_4}}^{*} \leq Z_{\mathrm{B,d,N_4}}^{\mathrm{alt},*} \leq \Big(W(0|0)\Big)^4.$$

Now we analyze the optimal value $Z_{\mathrm{B,p,N_4}}^{*}$ for the primal formulation, which is given in (102). If the belief $\boldsymbol{\beta}$ is as in (109), then we have

$$F_{\mathrm{B,p,N_4}}(\boldsymbol{\beta}) = -4 \log W(0|0),$$

which by the expression for $Z_{\mathrm{B,p,N_4}}^{*}$ in (102) implies

$$Z_{\mathrm{B,p,N_4}}^{*} \geq \Big(W(0|0)\Big)^4.$$

Then we have

$$Z_{\mathrm{B,p,N_4}}^{*} = Z_{\mathrm{B,d,N_4}}^{*} = Z_{\mathrm{B,d,N_4}}^{\mathrm{alt},*} = \Big(W(0|0)\Big)^4.$$

With this, we know that the beliefs in (109) correspond to the locations of the optimal value $Z_{\mathrm{B,p,N_4}}^{*}$. ■

Proposition 48. *Consider the S-NFG* N_4 *specified in Definition 44 under Assumption 45. One of the locations for the optimal value* $Z_{\mathrm{B,p,N_4}}^{*}$ *corresponds to an SPA fixed point whose messages are given by*

$$\boldsymbol{\mu}_{e\to f} = \Big(1, \quad 0\Big), \qquad e \in \partial f,\, f \in \mathcal{F}.$$

Moreover, the beliefs evaluated at this SPA fixed point are given by (109)*. The SPA is specified in Definition 5.*

Proof. It can be proven directly following the definition of the SPA in Definition 5, Assumption 45, and Theorem 47. Thus the details are omitted here. ■

Definition 49. *Based on Theorem 47, we make the following definitions.*

1) *Consider the belief* $\boldsymbol{\beta}$ *given in* (109)*. For each* $e \in \mathcal{E}$*, we define the set* $\mathcal{S}_e$ *to be*

$$\mathcal{S}_e \triangleq \Big\{ x_e \in \{0,1\} \;\Big|\; \beta_e(x_e) = 0 \Big\} = \{1\}.$$

The complement of $\mathcal{S}_e$ *is defined to be*

$$\mathcal{S}_e^{\mathrm{c}} \triangleq \mathcal{X}_e \setminus \mathcal{S}_e = \{0\}.$$

2) *For each function node $f \in \mathcal{F}$, we define the associated function f' to be*

$$f'(\boldsymbol{x}_f) \triangleq f(\boldsymbol{x}_f) \prod_{e\in\partial f} [x_e \in \mathcal{S}_e^{\mathrm{c}}] = \begin{cases} f(\mathbf{0}) & \boldsymbol{x}_f = \mathbf{0} \\ 0 & \textit{Otherwise} \end{cases}, \qquad \boldsymbol{x}_f \in \mathcal{X}_f. \tag{110}$$

3) *We define N_4' to be the factor graph consisting of the same vertex set $\mathcal{F}$, edge set $\mathcal{E}$, and alphabet $\mathcal{X}$ as N_4. However, for each vertex $f \in \mathcal{F}$, the associated local function is f' instead of f.*

Proposition 50. *Consider the S-NFG N_4 specified in Definition 44 under Assumption 45. The associated S-NFG specified in Definition 49 has the following properties.*

1) *The S-NFG N_4' has only one SPA fixed point, which is the same as the SPA fixed point in Proposition 48, and the beliefs evaluated at the SPA fixed point equal the beliefs in* (109). *Recall that the SPA is specified in Definition 5.*
2) $Z^*_{\mathrm{B,p},\mathsf{N}_4'} = Z^*_{\mathrm{B,d},\mathsf{N}_4'} = Z^{\mathrm{alt},*}_{\mathrm{B,d},\mathsf{N}_4'} = \Big(W(0|0)\Big)^4$.
3) *One of the locations of optimal value $Z^*_{\mathrm{B,p},\mathsf{N}_4'}$ for the primal formulation is given in* (109).

Proof. From the definition of f' in (110), we know that there is only one configuration of $(x_e)_{e\in\mathcal{E}}$ such that $\prod_f f'(\boldsymbol{x}_f) \neq 0$, which is $x_e = 0$ for all $e \in \mathcal{E}$. Note that when $x_e = 0$ for all $e \in \mathcal{E}$, we have $\prod_f f'(\boldsymbol{x}_f) = \Big(W(0|0)\Big)^4$. Thus the above properties can be proven straightforwardly and thus the details are omitted here. ■

VII. The Analysis of General S-NFGs

In this section, we consider the S-NFGs N satisfying Assumption 4 and relate the locations of the optimal value $F^{\mathrm{alt},*}_{\mathrm{B,d},\mathsf{N}}$ for the optimization problem in (22) to SPA fixed points. The results in this section generalize the results in Section IV. Here is the outline of this section.

1) The first main result will be presented in Theorem 54:

$$\lim_{m,n,k\to\infty} -\log\left(\prod_f Z_f\Big(\boldsymbol{\gamma}^{(m)}_{\partial f},\, \alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)}_{\partial f}\Big)\right) = F^{\mathrm{alt},*}_{\mathrm{B,d},\mathsf{N}}, \tag{111}$$

where Z_f is defined in (16).

2) The second main result is that the sequence $\{\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)}\}_{n,k}$ converges to a stationary point of $F^{\mathrm{alt}}_{\mathrm{B,d},\mathsf{N}}(\boldsymbol{\gamma},\boldsymbol{\lambda})$ w.r.t. $\boldsymbol{\lambda}$, i.e.,

$$\lim_{n,k\to\infty} \frac{\partial}{\partial\lambda_e(x_e)} F^{\mathrm{alt}}_{\mathrm{B,d},\mathsf{N}}\bigg|_{\boldsymbol{\lambda}=\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)}} = \lim_{n,k\to\infty}\Big(-b^{(m,n,k)}_{f_i,e}(x_e) + b^{(m,n,k)}_{f_j,e}(x_e)\Big) = 0, \qquad x_e \in \mathcal{X}_e,\ e=(f_i,f_j)\in\mathcal{E}, \tag{112}$$

which will be stated in Lemma 56. By the definition of $\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)}$ in Items 3 and 4 in Definition 17, the above expression obviously holds if the supremum of $F^{\mathrm{alt}}_{\mathrm{B,d},\mathsf{N}}$ w.r.t. $\boldsymbol{\lambda}$ is obtained in $\mathbb{R}^{|\mathcal{X}|}$. If the supremum is obtained in $\{\mathbb{R}\cup\{-\infty,+\infty\}\}^{|\mathcal{X}|}$, we need more details of $F^{\mathrm{alt}}_{\mathrm{B,d},\mathsf{N}}$. Because of

$$\frac{\partial}{\partial\lambda_e(x_e)} F^{\mathrm{alt}}_{\mathrm{B,d},\mathsf{N}} = -b_{f_i,e}(x_e) + b_{f_j,e}(x_e),$$

and the expressions in (203) and (204), we know that the entries in the gradient and the Hessian of $F^{\mathrm{alt}}_{\mathrm{B,d},\mathsf{N}}$ w.r.t. $\boldsymbol{\lambda}$ are finite. Thus we can analyze the second-order Taylor series expansion of $F^{\mathrm{alt}}_{\mathrm{B,d},\mathsf{N}}$ w.r.t. $\boldsymbol{\lambda}$, which is also finite when some entries in $\boldsymbol{\lambda}^{(n)}$ go to infinity.

3) The third main result will be presented in Theorem 57:

$$F^*_{\mathrm{B,p},\mathsf{N}} = F^*_{\mathrm{B,d},\mathsf{N}} = F^{\mathrm{alt},*}_{\mathrm{B,d},\mathsf{N}}.$$

The main idea of the proof is given as follows.

a) From the definition of $\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}$ in (37), we know that each entry in $\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}$ is bounded and there exists a subsequence $\left\{\boldsymbol{b}_{\mathcal{F}}^{(m,n_1,k_1)}\right\}_{n_1,k_1}$ such that all entries in $\boldsymbol{b}_{\mathcal{F}}^{(m,n_1,k_1)}$ converge for all $m \in \mathbb{Z}_{>0}$ as n_1 and k_1 go to infinity. By (112), we know that

$$\left(\lim_{n_1,k_1\to\infty} \boldsymbol{b}_{\mathcal{F}}^{(m,n_1,k_1)}\right) \in \mathcal{B}_{\mathcal{F}}(\mathsf{N}),$$

which means that $\boldsymbol{b}_{\mathcal{F}}^{(m,n_1,k_1)}$ converges to an element in $\mathcal{B}_{\mathcal{F}}(\mathsf{N})$ as n_1 and k_1 go to infinity. The LMP $\mathcal{B}_{\mathcal{F}}(\mathsf{N})$ is defined in (9).

b) In this proof, we consider the following Bethe free energy function:

$$F_{\mathrm{B,p,\mathsf{N}}}^{(1)}(\boldsymbol{\beta}_{\mathcal{F}}) = \sum_{f}\sum_{\boldsymbol{x}_f} \beta_f(\boldsymbol{x}_f)\cdot\log\frac{\beta_f(\boldsymbol{x}_f)}{f(\boldsymbol{x}_f)} - \sum_{e=(f_i,f_j)\in\mathcal{E}}\sum_{x_e}\frac{\beta_{f_i,e}(x_e)+\beta_{f_j,e}(x_e)}{2}\cdot\log\left(\frac{\beta_{f_i,e}(x_e)+\beta_{f_j,e}(x_e)}{2}\right),$$

which is defined in (11). As shown in the definition of $F_{\mathrm{B,p,\mathsf{N}}}^{(1)}$, we do not require $\boldsymbol{\beta}_{\mathcal{F}} \in \mathcal{B}_{\mathcal{F}}(\mathsf{N})$ here. By setting

$$\boldsymbol{\beta}_f = \boldsymbol{b}_f^{(m,n,k)}, \qquad f \in \mathcal{F},$$

we have

$$\begin{aligned} F_{\mathrm{B,p,\mathsf{N}}}^{(1)}\left(\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}\right) &= -\log\left(\prod_f Z_f\big(\boldsymbol{\gamma}_{\partial f}^{(m)},\, \alpha^{(k)}\cdot\boldsymbol{\lambda}_{\partial f}^{(n)}\big)\right) + \sum_f\sum_{\boldsymbol{x}_f} b_f^{(m,n,k)}(\boldsymbol{x}_f)\cdot\sum_{e\in\partial f}\alpha^{(k)}\cdot\lambda_{e,f}^{(n)}(x_{e,f}) \\ &\quad + \sum_{e=(f_i,f_j)\in\mathcal{E}}\ \sum_{x_e:\,\gamma_e^{(m)}(x_e)>0}\frac{b_{f_i,e}^{(m,n,k)}(x_e)+b_{f_j,e}^{(m,n,k)}(x_e)}{2}\cdot\log\left(\frac{2\gamma_e^{(m)}(x_e)}{b_{f_i,e}^{(m,n,k)}(x_e)+b_{f_j,e}^{(m,n,k)}(x_e)}\right). \end{aligned} \tag{113}$$

Then we have

$$\lim_{n_1,k_1\to\infty} F_{\mathrm{B,p,\mathsf{N}}}^{(1)}\left(\boldsymbol{b}_{\mathcal{F}}^{(m,n_1,k_1)}\right) \overset{(a)}{\geq} F_{\mathrm{B,p,\mathsf{N}}}^{*} \overset{(b)}{=} F_{\mathrm{B,d,\mathsf{N}}}^{*}, \qquad m\in\mathbb{Z}_{>0}. \tag{114}$$

where step (a) follows from the continuity of $F_{\mathrm{B,p,\mathsf{N}}}^{(1)}$ w.r.t. $\boldsymbol{\beta}_{\mathcal{F}}$ and the definitions of $F_{\mathrm{B,p,\mathsf{N}}}^{*}$ in (13) and where step (b) follows from Proposition 16. Then, in order to prove that the above inequality is indeed an equality, we first need to show that

$$\lim_{n,k\to\infty}\sum_f\sum_{\boldsymbol{x}_f} b_f^{(m,n_1,k_1)}(\boldsymbol{x}_f)\cdot\sum_{e\in\partial f}\alpha^{(k)}\cdot\lambda_{e,f}^{(n)}(x_{e,f}) = 0, \tag{115}$$

which will be proven in Lemma 63. The main idea of the proof of Lemma 63 is as follows.

- To have (115), we need to define $\alpha^{(k)}$ following the way in Item 4 in Definition 17.
- Then we prove that $\alpha^{(k)}$ converges to the stationary point of $F_{\mathrm{B,d,\mathsf{N}}}^{\mathrm{alt}}(\boldsymbol{\gamma},\alpha\cdot\boldsymbol{\lambda})$ w.r.t. α, which proves (115).

From this, it follows that

$$\begin{aligned} &\limsup_{m\to\infty}\lim_{n_1,k_1\to\infty} F_{\mathrm{B,p,\mathsf{N}}}^{(1)}\left(\boldsymbol{b}_{\mathcal{F}}^{(m,n_1,k_1)}\right) \\ &\leq \limsup_{m,n,k\to\infty} F_{\mathrm{B,p,\mathsf{N}}}^{(1)}\left(\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}\right) \\ &\leq -\liminf_{m,n,k\to\infty}\log\left(\prod_f Z_f\big(\boldsymbol{\gamma}_{\partial f}^{(m)},\,\alpha^{(k)}\cdot\boldsymbol{\lambda}_{\partial f}^{(n)}\big)\right) + \limsup_{m,n,k\to\infty}\sum_f\sum_{\boldsymbol{x}_f} b_f^{(m,n,k)}(\boldsymbol{x}_f)\cdot\sum_{e\in\partial f}\alpha^{(k)}\cdot\lambda_{e,f}^{(n)}(x_{e,f}) \\ &\quad + \limsup_{m,n,k\to\infty}\sum_{e=(f_i,f_j)\in\mathcal{E}}\ \sum_{x_e:\,\gamma_e^{(m)}(x_e)>0}\frac{b_{f_i,e}^{(m,n,k)}(x_e)+b_{f_j,e}^{(m,n,k)}(x_e)}{2}\cdot\log\left(\frac{2\gamma_e^{(m)}(x_e)}{b_{f_i,e}^{(m,n,k)}(x_e)+b_{f_j,e}^{(m,n,k)}(x_e)}\right) \end{aligned}$$

$$\stackrel{(a)}{=} F_{\mathrm{B,d,N}}^{\mathrm{alt},*} + \limsup_{m,n,k\to\infty} \sum_{e=(f_i,f_j)\in\mathcal{E}} \sum_{x_e:\, \gamma_e^{(m)}(x_e)>0} \frac{b_{f_i,e}^{(m,n,k)}(x_e) + b_{f_j,e}^{(m,n,k)}(x_e)}{2} \cdot \log\left(\frac{2\gamma_e^{(m)}(x_e)}{b_{f_i,e}^{(m,n,k)}(x_e) + b_{f_j,e}^{(m,n,k)}(x_e)} \right), \tag{116}$$

where step (a) follows from (111) and (115). Using the fact that the Kullback–Leibler (K-L) divergence is nonnegative, we have

$$\limsup_{m\to\infty} \lim_{n_1,k_1\to\infty} F_{\mathrm{B,p,N}}^{(1)}\left(\boldsymbol{b}_{\mathcal{F}}^{(m,n_1,k_1)}\right) \leq F_{\mathrm{B,d,N}}^{\mathrm{alt},*} \stackrel{(a)}{\leq} F_{\mathrm{B,d,N}}^{*},$$

where step (a) follows from Proposition 15. Combining with (114), we have

$$\limsup_{m\to\infty} \lim_{n_1,k_1\to\infty} F_{\mathrm{B,p,N}}^{(1)}\left(\boldsymbol{b}_{\mathcal{F}}^{(m,n_1,k_1)}\right) = F_{\mathrm{B,d,N}}^{\mathrm{alt},*} = F_{\mathrm{B,d,N}}^{*} = F_{\mathrm{B,p,N}}^{*}.$$

4) By (116) and the third main result, we obtain the forth main result

$$\lim_{m,n_1,k_1\to\infty} \sum_{e=(f_i,f_j)\in\mathcal{E}} \sum_{x_e:\, \gamma_e^{(m)}(x_e)>0} \frac{b_{f_i,e}^{(m,n_1,k_1)}(x_e) + b_{f_j,e}^{(m,n_1,k_1)}(x_e)}{2} \cdot \log\left(\frac{2\gamma_e^{(m)}(x_e)}{b_{f_i,e}^{(m,n_1,k_1)}(x_e) + b_{f_j,e}^{(m,n_1,k_1)}(x_e)} \right) = 0.$$

By Pinsker's inequality (see, e.g., [20, Theorem 2.33].), we have

$$\lim_{m,n_1,k_1\to\infty} \sum_{e=(f_i,f_j)\in\mathcal{E}} \sum_{x_e:\, \gamma_e^{(m)}(x_e)>0} \left| \frac{b_{f_i,e}^{(m,n_1,k_1)}(x_e) + b_{f_j,e}^{(m,n_1,k_1)}(x_e)}{2} - \gamma_e^{(m)}(x_e) \right| = 0.$$

Because the entries in $\boldsymbol{b}_{\mathcal{F}}^{(m,n_1,k_1)}$ and $\boldsymbol{\gamma}^{(m)}$ are bounded, there exists a subsequence $\{m_1\}$ of $\{m\}$ such that all entries of $\{\boldsymbol{b}_{\mathcal{F}}^{(m_1,n_1,k_1)}\}_{m_1\in\mathbb{Z}_{>0}}$ and $\{\boldsymbol{\gamma}^{(m_1)}\}_{m_1\in\mathbb{Z}_{>0}}$ converge. Combining the above equality with (112), we have

$$\lim_{m_1,n_1,k_1\to\infty} b_{f_i,e}^{(m_1,n_1,k_1)}(x_e) = \lim_{m_1,n_1,k_1\to\infty} b_{f_j,e}^{(m_1,n_1,k_1)}(x_e) = \lim_{m_1\to\infty} \gamma_e^{(m_1)}(x_e), \qquad x_e \in \mathcal{X}_e,\, e=(f_i,f_j)\in\mathcal{E}.$$

This result will be stated in (132) in Theorem 57.

5) The last main result will be presented in Theorem 61. We will show that there exists a sequence $\{\boldsymbol{\gamma}^{(m_2)}, \boldsymbol{\lambda}^{(n_2)}, \alpha^{(k_2)}\}_{m_2,n_2,k_2\in\mathbb{Z}_{>0}}$ such that the collection of messages $\boldsymbol{\mu}^{(m_2,n_2,k_2)}$, which is defined based on $\boldsymbol{\gamma}^{(m_2)}$, $\boldsymbol{\lambda}^{(n_2)}$, and $\alpha^{(k_2)}$, converges to SPA fixed-point messages for a modified S-NFG N' as m_2, n_2, k_2 go to infinity. The details of N' will be given in Definition 58. The proof technique is standard.

Before proving the main results in this paper, we wonder whether the functions specified in Definition 17 for the dual formulation is well defined. We answer this question by proving the following properties.

1) The optimal value $F_{\mathrm{B,d,N}}^{\mathrm{alt},*}$ of the optimization problem in (21) is real-valued.
2) It is sufficient to consider a sequence $\boldsymbol{\gamma}^{(m)}$ such that the sequence $\left\{\hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}\left(\boldsymbol{\gamma}^{(m)}\right)\right\}_m$ converges to $F_{\mathrm{B,d,N}}^{\mathrm{alt},*}$ as defined in (30) and $\hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}\left(\boldsymbol{\gamma}^{(m)}\right)$ is real-valued for each $m\in\mathbb{Z}_{>0}$.
3) For each m, n, and k, the function $Z_f\big(\boldsymbol{\gamma}_{\partial f}^{(m)},\, \alpha^{(k)}\cdot\boldsymbol{\lambda}_{\partial f}^{(n)}\big)$ is positive-valued.
4) Based on the previous property, the belief $b_f^{(m,n,k)}(\boldsymbol{x}_f)$ in (37) for $f\in\mathcal{F}$, which is defined based on the sequences $\boldsymbol{\gamma}_{\partial f}^{(m)}$, $\boldsymbol{\lambda}_{\partial f}^{(n)}$, and $\alpha^{(k)}$, is nonnegative real-valued.

Proposition 51. *Because the considered S-NFG N satisfies Assumption 4, then*

$$\exists \boldsymbol{x}' = (x'_e)_{e\in\mathcal{E}} \in \mathcal{X} \text{ such that } g(\boldsymbol{x}') > 0. \tag{117}$$

It holds that

$$-\sum_f \log\left(\sum_{\boldsymbol{x}_{\partial f,f}} f(\boldsymbol{x}_{\partial f,f}) \right) \leq -\log Z_{\mathrm{B,d,N}}^{\mathrm{alt},*} = F_{\mathrm{B,d,N}}^{\mathrm{alt},*} \leq -\log g(\boldsymbol{x}') < \infty, \tag{118}$$

$$-\sum_{f}\log\left(\sum_{\boldsymbol{x}_{\partial f,f}} f(\boldsymbol{x}_{\partial f,f})\right) \leq F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)},\mathbf{0}) \leq \hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)}\Big), \qquad m\in\mathbb{Z}_{>0}. \tag{119}$$

Proof. See Appendix N. ■

Proposition 52. *We present some properties on the functions specified in Definitions 12 and 17. Some of the properties are stated in the definition. Here we make a summary of them.*

1) *The function $F_{\mathrm{B,d,N}}^{\mathrm{alt}}$ evaluated at $(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)})$ is given by*

$$\begin{aligned}F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)}\Big) &= -\log\left(\prod_{f} Z_f\Big(\boldsymbol{\gamma}_{\partial f}^{(m)},\,\alpha^{(k)}\cdot\boldsymbol{\lambda}_{\partial f}^{(n)}\Big)\right)\\ &= -\log\left(\prod_{f\in\mathcal{F}}\left(\sum_{\boldsymbol{x}_{\partial f,f}} f(\boldsymbol{x}_{\partial f,f})\cdot\prod_{e\in\partial f}\Big(\exp\big(\alpha^{(k)}\cdot\lambda_{e,f}^{(n)}(x_{e,f})\big)\cdot\sqrt{\gamma_e^{(m)}(x_{e,f})}\Big)\right)\right).\end{aligned} \tag{120}$$

2) *It is sufficient to consider a sequence $\{\boldsymbol{\gamma}^{(m)}\}_m$ such that $\hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)}\Big)$ is bounded:*

$$F_{\mathrm{B,d,N}}^{\mathrm{alt},*} \leq \hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)}\Big) \leq -\log g(\boldsymbol{x}'), \qquad \forall m\in\mathbb{Z}_{>0}. \tag{121}$$

3) *The sequences $\{\boldsymbol{\gamma}^{(m)}\}_m$, $\{\boldsymbol{\lambda}^{(n)}\}_n$ and $\{\alpha^{(k)}\}_k$ have the following properties*
   a) *The function Z_f evaluated at these sequences is positive-valued:*

$$Z_f\Big(\boldsymbol{\gamma}_{\partial f}^{(m)},\,\alpha^{(k)}\cdot\boldsymbol{\lambda}_{\partial f}^{(n)}\Big)\in\mathbb{R}_{>0}, \qquad m,n,k\in\mathbb{Z}_{>0},\ f\in\mathcal{F}. \tag{122}$$

   b) *With* (122)*, there exists an $\boldsymbol{x}_f\in\mathcal{X}_f$ such that*

$$b_f^{(m,n,k)}(\boldsymbol{x}_f)\in\mathbb{R}_{>0}, \qquad m,n,k\in\mathbb{Z}_{>0}.$$

   c) *The function $F_{\mathrm{B,d,N}}^{\mathrm{alt}}$ evaluated at $(\boldsymbol{\gamma}^{(m)},\mathbf{0})$ is also bounded:*

$$-\sum_{f}\log\left(\sum_{\boldsymbol{x}_{\partial f,f}} f(\boldsymbol{x}_{\partial f,f})\right) \leq F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)},\mathbf{0}) \leq -\log g(\boldsymbol{x}'), \qquad m\in\mathbb{Z}_{>0}, \tag{123}$$

   where $g(\boldsymbol{x}')$ is given in (117)*.*
   d) *The sequence $\Big\{\boldsymbol{\gamma}^{(m)}\Big\}_m$ satisfies*

$$\lim_{m\to\infty}\hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)}\Big) = F_{\mathrm{B,d,N}}^{\mathrm{alt},*}\in\mathbb{R}. \tag{124}$$

   e) *For fixed $\boldsymbol{\gamma}^{(m)}$, it hold that*

$$\lim_{n\to\infty}F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)},\boldsymbol{\lambda}^{(n)}\Big) = \sup_{\boldsymbol{\lambda}}F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)},\boldsymbol{\lambda}\Big) = \hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)}\Big)\in\mathbb{R}. \tag{125}$$

   f) *For Fixed $\boldsymbol{\gamma}^{(m)}$ and $\boldsymbol{\lambda}^{(n)}$, it holds that*

$$\lim_{k\to\infty}F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)}\Big) = \max_{\alpha\in\mathbb{R}\cup\{-\infty,+\infty\}}F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)},\alpha\cdot\boldsymbol{\lambda}^{(n)}\Big)\in\mathbb{R}. \tag{126}$$

Proof. See Appendix O. ■

Now we move on to the proof of the first main result. From (30) and (32), we know that

$$\lim_{m,n\to\infty}F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)},\boldsymbol{\lambda}^{(n)}\Big) = F_{\mathrm{B,d,N}}^{\mathrm{alt},*}.$$

However, if we replace $\boldsymbol{\lambda}^{(n)}$ with $\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)}$, then we want to know whether the sequence $\Big\{F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)}\Big)\Big\}_{m,n,k}$ converges. To be more specific, we want to prove that

$$\lim_{m,n,k\to\infty}F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)}\Big) = F_{\mathrm{B,d,N}}^{\mathrm{alt},*}.$$

To begin with, we prove that the associated sequence is a bounded sequence.

Proposition 53. *For each $\boldsymbol{\gamma}^{(m)}$, the following sequence is bounded:*

$$\left\{ F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}) \right\}_{n,k}. \tag{127}$$

The following inequality also holds

$$\prod_{f \in \mathcal{F}} \left(f(\boldsymbol{x}_{\partial f,f}) \cdot \prod_{e \in \partial f} \left(\exp\bigl(\alpha^{(k)} \cdot \lambda_{e,f}^{(n)}(x_{e,f})\bigr) \cdot \sqrt{\gamma_e^{(m)}(x_{e,f})} \right) \right) \leq \exp\Bigl(-F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}, \mathbf{0})\Bigr) \in \mathbb{R}, \qquad m \in \mathbb{Z}_{>0}. \tag{128}$$

Proof. By (125), the sequence $\{\boldsymbol{\lambda}^{(n)}\}$ satisfies

$$\hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}\Bigl(\boldsymbol{\gamma}^{(m)}\Bigr) = \lim_{n \to \infty} F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}, \boldsymbol{\lambda}^{(n)}) = \sup_{\boldsymbol{\lambda} \in \mathbb{R}^{|\mathcal{X}|}} F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}, \boldsymbol{\lambda}) \in \mathbb{R},$$

where $\hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}$ is defined in (29). By (126), the sequence $\{\alpha^{(k)}\}_k$ satisfies

$$\lim_{k \to \infty} F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Bigl(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}\Bigr) = \sup_{\alpha \in \mathbb{R}} F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Bigl(\boldsymbol{\gamma}^{(m)}, \alpha \cdot \boldsymbol{\lambda}^{(n)}\Bigr) \in \mathbb{R}, \qquad \boldsymbol{\lambda}^{(n)} \in \mathbb{R}^{|\mathcal{X}|}.$$

Thus

$$F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Bigl(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}\Bigr) \leq \hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}\Bigl(\boldsymbol{\gamma}^{(m)}\Bigr) \overset{(a)}{\in} \mathbb{R}, \qquad n, k \in \mathbb{Z}_{>0},$$

where step (a) follows from the inequalities (121) in Proposition 52. Then we have

$$F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Bigl(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}\Bigr) \overset{(a)}{\geq} F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}, \mathbf{0}) \overset{(b)}{\in} \mathbb{R}, \qquad n, k \in \mathbb{Z}_{>0},$$

where step (a) follows from the definition of $\{\alpha^{(k)}\}$ in (34) and (35) and where step (b) follows from the property of $F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}, \mathbf{0})$ in (123). Thus the sequence in (127) is bounded. By (120), the above inequality implies that

$$\begin{aligned} \prod_f Z_f\bigl(\boldsymbol{\gamma}_{\partial f}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}_{\partial f}^{(n)}\bigr) &= \prod_{f \in \mathcal{F}} \left(\sum_{\boldsymbol{x}_{\partial f,f}} f(\boldsymbol{x}_{\partial f,f}) \cdot \prod_{e \in \partial f} \left(\exp\bigl(\alpha^{(k)} \cdot \lambda_{e,f}^{(n)}(x_{e,f})\bigr) \cdot \sqrt{\gamma_e^{(m)}(x_{e,f})} \right) \right) \\ &\leq \exp\Bigl(-F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}, \mathbf{0})\Bigr) \\ &\in \mathbb{R}, \qquad m \in \mathbb{Z}_{>0}, \end{aligned}$$

which proves (128). ■

Theorem 54. *With fixed $\boldsymbol{\gamma}^{(m)}$, it holds that:*

$$\begin{aligned} \lim_{n,k \to \infty} F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Bigl(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}\Bigr) &= \lim_{n,k \to \infty} -\log \prod_f Z_f\bigl(\boldsymbol{\gamma}_{\partial f}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}_{\partial f}^{(n)}\bigr) \\ &= \hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}\Bigl(\boldsymbol{\gamma}^{(m)}\Bigr) \\ &\in \mathbb{R}, \qquad m \in \mathbb{Z}_{>0}. \end{aligned}$$

Proof. See Appendix P. ■

Corollary 55. *It holds that*

$$\lim_{m,n,k \to \infty} F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Bigl(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}\Bigr) = F_{\mathrm{B,d,N}}^{\mathrm{alt},*}.$$

Proof. It can be proven straightforwardly by Theorem 54 and the definition of $\{\boldsymbol{\gamma}^{(m)}\}$ in (30). ■

It is natural to ask whether $F_{\mathrm{B,d,N}}^{\mathrm{alt},*}$ and $F_{\mathrm{B,d,N}}^{*}$ are equal. By Propositions 15 and 16, we know that

$$F_{\mathrm{B,d,N}}^{\mathrm{alt},*} \leq F_{\mathrm{B,d,N}}^{*} = F_{\mathrm{B,p,N}}^{*}.$$

In the following, we show $F_{\mathrm{B,d,N}}^{\mathrm{alt},*} = F_{\mathrm{B,p,N}}^{*}$ by analyzing the function $F_{\mathrm{B,p,N}}^{(1)}$ for the primal formulation as specified in Definition 8. The main idea is listed as follows.

1) We consider the collection of beliefs $\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}$, which is defined based on the sequences $\{\boldsymbol{\gamma}^{(m)}\}_m$, $\{\boldsymbol{\lambda}^{(n)}\}_n$, and $\{\alpha^{(k)}\}_k$. We want to show that $F_{\mathrm{B,p,N}}^{(1)}(\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)})$ converges to $F_{\mathrm{B,p,N}}^{*}$ as m, n, and k go to infinity.
2) By the definition of $F_{\mathrm{B,p,N}}^{*}$ in (13) and Proposition 9, we know that it is related to the minimum of the alternative Bethe free energy function $F_{\mathrm{B,p,N}}^{(1)}$ over a constrained set $\mathcal{B}_{\mathcal{F}}(\mathsf{N})$. Thus we need to show that $\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}$ converges to an element in $\mathcal{B}_{\mathcal{F}}(\mathsf{N})$. By the definition of $\mathcal{B}_{\mathcal{F}}(\mathsf{N})$ in (9), it means that we need to show that the elements in $\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}$ satisfy the local consistency constraints as m, n, and k go to infinity.
3) In the expression of $F_{\mathrm{B,p,N}}^{(1)}(\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)})$, we note that it consists of $F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}\Big)$ and other terms. By Corollary 55, we know that $F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}\Big)$ converges to $F_{\mathrm{B,d,N}}^{\mathrm{alt},*}$. If we can show that the remaining terms converge to some non-positive terms, then we have proven that $F_{\mathrm{B,p,N}}^{*} \leq F_{\mathrm{B,d,N}}^{\mathrm{alt}}$.

Lemma 56. *It holds that*

$$\lim_{n,k\to\infty} \Big(b_{f_j,e}^{(m,n,k)}(x_e) - b_{f_i,e}^{(m,n,k)}(x_e) \Big) = 0, \qquad \boldsymbol{\gamma} = \boldsymbol{\gamma}^{(m)},\ x_e \in \mathcal{X}_e,\ e \in \mathcal{E},$$

where $\boldsymbol{b}_{f,e}^{(m,n,k)}$ *is defined in* (39).

Proof. See Appendix Q. ■

By setting

$$\boldsymbol{\beta}_f = \boldsymbol{b}_f^{(m,n,k)}, \qquad f \in \mathcal{F},$$

where $\boldsymbol{b}_f^{(m,n,k)}$ is defined in (37), the function $F_{\mathrm{B,p,N}}^{(1)}$ defined in (11) equals

$$\begin{aligned} F_{\mathrm{B,p,N}}^{(1)}\Big(\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}\Big) &= -\log\left(\prod_f Z_f\big(\boldsymbol{\gamma}_{\partial f}^{(m)},\, \alpha^{(k)} \cdot \boldsymbol{\lambda}_{\partial f}^{(n)}\big) \right) + \sum_f \sum_{\boldsymbol{x}_f} b_f^{(m,n,k)}(\boldsymbol{x}_f) \cdot \sum_{e\in\partial f} \alpha^{(k)} \cdot \lambda_{e,f}^{(n)}(x_{e,f}) \\ &\quad + \sum_{e=(f_i,f_j)\in\mathcal{E}} \ \sum_{x_e:\, \gamma_e^{(m)}(x_e)>0} \frac{b_{f_i,e}^{(m,n,k)}(x_e) + b_{f_j,e}^{(m,n,k)}(x_e)}{2} \cdot \log\left(\frac{2\gamma_e^{(m)}(x_e)}{b_{f_i,e}^{(m,n,k)}(x_e) + b_{f_j,e}^{(m,n,k)}(x_e)} \right), \end{aligned} \tag{129}$$

where $\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)} = \Big(\boldsymbol{b}_f^{(m,n,k)}\Big)_f$ as defined in (38).

Theorem 57. *The function* $F_{\mathrm{B,p,N}}^{(1)}$ *has the following property:*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \limsup_{k\to\infty} F_{\mathrm{B,p,N}}^{(1)}\Big(\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}\Big) = F_{\mathrm{B,d,N}}^{\mathrm{alt},*} = F_{\mathrm{B,d,N}}^{*} = F_{\mathrm{B,p,N}}^{*}, \tag{130}$$

where $\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}$ *is given in* (38)*, the function* $F_{\mathrm{B,p,N}}^{(1)}$ *is given in* (129)*, the quantity* $F_{\mathrm{B,d,N}}^{\mathrm{alt},*}$ *is defined in* (22)*, the quantity* $F_{\mathrm{B,d,N}}^{*}$ *is defined in* (20)*, and the quantity* $F_{\mathrm{B,p,N}}^{*}$ *is the minimum of the constrained Bethe partition function defined in* (13)*. There exists subsequences of* $\{\boldsymbol{\gamma}^{(m)}\}_m$*,* $\{\boldsymbol{\lambda}^{(n)}\}_n$*, and* $\{\alpha^{(k)}\}_k$ *indexed by* $\{m_1, n_1, k_1\}$ *such that*

$$\Big(\lim_{m_1,n_1,k_1\to\infty} \boldsymbol{b}_{\mathcal{F}}^{(m_1,n_1,k_1)} \Big) \in \mathcal{B}_{\mathcal{F}}(\mathsf{N}), \tag{131}$$

$$\lim_{m_1,n_1,k_1\to\infty} b_{f_i,e}^{(m_1,n_1,k_1)}(x_e) = \lim_{m_1,n_1,k_1\to\infty} b_{f_j,e}^{(m_1,n_1,k_1)}(x_e) = \lim_{m_1\to\infty} \gamma_e^{(m_1)}(x_e), \quad x_e \in \mathcal{X}_e,\ e = (f_i, f_j) \in \mathcal{E}, \tag{132}$$

$$\lim_{m_1,n_1,k_1\to\infty} F_{\mathrm{B,p,N}}^{(1)}(\boldsymbol{b}_{\mathcal{F}}^{(m_1,n_1,k_1)}) = F_{\mathrm{B,d,N}}^{\mathrm{alt},*} = F_{\mathrm{B,d,N}}^{*} = F_{\mathrm{B,p,N}}^{*}, \tag{133}$$

where $\mathcal{B}_{\mathcal{F}}(\mathsf{N})$ *is defined in* (9).

Proof. See Appendix R. ■

As mentioned in the introduction section, we want to relate location to the global minimum of the Bethe free energy function to an SPA fixed point. In Theorem 57, we know that $\left\{\boldsymbol{b}_{\mathcal{F}}^{(m_1,n_1,k_1)}\right\}_{m_1,n_1,k_1}$ converges to one of the locations of the minimum of the Bethe free energy function. We want to know that whether there is a collection of messages defined based on $\boldsymbol{\gamma}^{(m_1)}$, $\boldsymbol{\lambda}^{(n_1)}$, and $\alpha^{(k_1)}$ such that this collection of messages converges to SPA fixed-point messages.

Definition 58. *Based on Theorem 57, we make the following definitions.*

1) *We define $\boldsymbol{b}_{\mathcal{F}}^*$ to be*

$$\boldsymbol{b}_{\mathcal{F}}^* \triangleq \{\boldsymbol{b}_f^*\}, \qquad b_f^*(\boldsymbol{x}_f) \triangleq \lim_{m_1,n_1,k_1\to\infty} b_f^{(m_1,n_1,k_1)}(\boldsymbol{x}_f), \qquad \boldsymbol{x}_f \in \mathcal{X}_f,\, f \in \mathcal{F}. \tag{134}$$

2) *We define $\boldsymbol{\gamma}^*$ to be $\boldsymbol{\gamma}^* \triangleq \{\gamma_e^*(x_e)\}_{x_e\in\mathcal{X}_e, e\in\mathcal{E}}$ with entries given by*

$$\gamma_e^*(x_e) \triangleq \lim_{m_1\to\infty} \gamma_e^{(m_1)}(x_e), \qquad x_e \in \mathcal{X}_e,\, e \in \mathcal{E}. \tag{135}$$

3) *For each $e \in \mathcal{E}$, we define the set $\mathcal{S}_e$ to be*

$$\mathcal{S}_e \triangleq \left\{x_e \in \mathcal{X}_e \,|\gamma_e^*(x_e) = 0\right\}.$$

The complement of $\mathcal{S}_e$ is defined to be

$$\mathcal{S}_e^c \triangleq \mathcal{X}_e \setminus \mathcal{S}_e. \tag{136}$$

4) *We define f' to be*

$$f'(\boldsymbol{x}_f) \triangleq f(\boldsymbol{x}_f) \cdot \prod_{e\in\partial f} [x_e \in \mathcal{S}_e^c], \qquad \boldsymbol{x}_f \in \mathcal{X}_f,\, f \in \mathcal{F}.$$

5) *We define N' to be the factor graph consisting of the same vertex set $\mathcal{F}$, edge set $\mathcal{E}$, and alphabet $\mathcal{X}$ as N. However, for each vertex f, the associated local function is f' instead of f.*

Proposition 59. *The collection of vector $\boldsymbol{b}_{\mathcal{F}}^*$ corresponds to one of the locations of the minimum of the Bethe free energy function $F_{\mathrm{B,p,N'}}^{(1)}(\boldsymbol{\beta}_{\mathcal{F}})$ over $\mathcal{B}_{\mathcal{F}}(\mathsf{N}')$, i.e.,*

$$F_{\mathrm{B,p,N}}^* = F_{\mathrm{B,p,N'}}^* = F_{\mathrm{B,p,N}}^{(1)}(\boldsymbol{b}_{\mathcal{F}}^*) = F_{\mathrm{B,p,N'}}^{(1)}(\boldsymbol{b}_{\mathcal{F}}^*),$$

where the function $F_{\mathrm{B,p,N}}^{(1)}$ is given in (129)*, and the quantity $F_{\mathrm{B,p,N}}^*$ is the minimum of the constrained Bethe partition function defined in* (13)*, and the set $\mathcal{B}_{\mathcal{F}}(\mathsf{N}')$ is defined in* (9)*.*

Proof. In this proof, we consider $F_{\mathrm{B,p,N}}^{(1)}$ defined in (11), which is an alternative form of $F_{\mathrm{B,p,N}}$ defined in (10). By (131) and (132) in Theorem 57 and the definition of $\boldsymbol{b}_f^*$ in (134), we know that

$$\boldsymbol{b}_{\mathcal{F}}^* \in \mathcal{B}_{\mathcal{F}}(\mathsf{N}), \qquad \boldsymbol{b}_{\mathcal{F}}^* \in \mathcal{B}_{\mathcal{F}}(\mathsf{N}').$$

It holds that

$$F_{\mathrm{B,p,N}}^* \overset{(a)}{=} \lim_{m_1,n_1,k_1\to\infty} F_{\mathrm{B,p,N}}^{(1)}\left(\boldsymbol{b}_{\mathcal{F}}^{(m_1,n_1,k_1)}\right) \overset{(b)}{=} F_{\mathrm{B,p,N}}^{(1)}(\boldsymbol{b}_{\mathcal{F}}^*) \overset{(c)}{=} F_{\mathrm{B,p,N'}}^{(1)}(\boldsymbol{b}_{\mathcal{F}}^*) \overset{(d)}{\geq} F_{\mathrm{B,p,N'}}^*,$$

where step (a) follows from (133) in Theorem 57, where step (b) follows from the continuity of $F_{\mathrm{B,p,N}}^{(1)}(\boldsymbol{\beta}_{\mathcal{F}})$ w.r.t. $\boldsymbol{\beta}_{\mathcal{F}}$, where step (c) follows from the definition of N' in Item 5 in Definition 58, and where step (d) follows from the definition of $F_{\mathrm{B,p,N'}}^*$ in (13). Denote the location of $F_{\mathrm{B,p,N'}}^*$ as

$$\{\boldsymbol{b}_{f'}^*\}_{f\in\mathcal{F}} \in \mathrm{argmin}_{\{\boldsymbol{b}_{f'}\}_{f\in\mathcal{F}}\in\mathcal{B}(\mathsf{N}')}\, F_{\mathrm{B,p,N'}}^{(1)}\big(\{\boldsymbol{b}_{f'}\}_{f\in\mathcal{F}}\big).$$

We have

$$b^*_{f',e}(x_e) = 0, \qquad x_e \in \mathcal{S}_e,\, e \in \partial f,\, f \in \mathcal{F},$$

which implies

$$F^*_{\mathrm{B,p},\mathsf{N}'} = F^{(1)}_{\mathrm{B,p},\mathsf{N}'}\big(\{\boldsymbol{b}^*_{f'}\}_f\big) = F^{(1)}_{\mathrm{B,p},\mathsf{N}}\big(\{\boldsymbol{b}^*_{f'}\}_f\big) \geq F_{\mathrm{B,p},\mathsf{N}}.$$

■

Definition 60. *In this definition, for consistency with Definition 5, with a slight abuse of notation, we define*

$$\mu^{(m,n,k)}_{e\to f}(x_e) \triangleq \begin{cases} \exp\big(\alpha^{(k)} \cdot \lambda^{(n)}_{e,f}(x_e)\big) \cdot \sqrt{\gamma^{(m)}_e(x_e)} & x_e \in \mathcal{S}^c_e \\ 0 & \textit{otherwise} \end{cases}. \tag{137}$$

$$Z^{(m,n,k)}_{\mu_{e\to f}} \triangleq \sum_{x_e \in \mathcal{S}^c_e} \mu^{(m,n,k)}_{e\to f}(x_e), \qquad e \in \partial f,\, f \in \mathcal{F}. \tag{138}$$

For each $e = (f_i, f_j) \in \mathcal{E}$, we define

$$\mu^{(m,n,k)}_{e\to f_i,\mathrm{SPA}}(x_e) \triangleq \begin{cases} \frac{1}{C^{(m,n,k)}_{e\to f_i}} \cdot \frac{b^{(m,n,k)}_{f_j,e}(x_e)}{\mu^{(m,n,k)}_{e\to f_j}(x_e)} & x_e \in \mathcal{S}^c_e \\ 0 & \textit{otherwise} \end{cases}, \quad \mu^{(m,n,k)}_{e\to f_j,\mathrm{SPA}}(x_e) \triangleq \begin{cases} \frac{1}{C^{(m,n,k)}_{e\to f_j}} \cdot \frac{b^{(m,n,k)}_{f_i,e}(x_e)}{\mu^{(m,n,k)}_{e\to f_i}(x_e)} & x_e \in \mathcal{S}^c_e \\ 0 & \textit{otherwise} \end{cases} \tag{139}$$

where normalization constants $C^{(m,n,k)}_{e\to f_i}$ and $C^{(m,n,k)}_{e\to f_j}$ are defined to be:

$$C^{(m,n,k)}_{e\to f_i} \triangleq \sum_{x_e \in \mathcal{S}^c_e} \frac{b^{(m,n,k)}_{f_j,e}(x_e)}{\mu^{(m,n,k)}_{e\to f_j}(x_e)}, \qquad C^{(m,n,k)}_{e\to f_j} \triangleq \sum_{x_e \in \mathcal{S}^c_e} \frac{b^{(m,n,k)}_{f_i,e}(x_e)}{\mu^{(m,n,k)}_{e\to f_i}(x_e)}. \tag{140}$$

For simplicity, we also define

$$\boldsymbol{\mu}^{(m,n,k)} \triangleq \Big(\mu^{(m,n,k)}_{e\to f}(x_e)\Big)_{x_e \in \mathcal{X}_e, e\in\partial f, f \in \mathcal{F}}, \qquad \boldsymbol{\mu}^{(m,n,k)}_{\mathrm{SPA}} \triangleq \Big(\mu^{(m,n,k)}_{e\to f,\mathrm{SPA}}(x_e)\Big)_{x_e \in \mathcal{X}_e, e\in\partial f, f \in \mathcal{F}}.$$

Theorem 61. *There exists a subsequence of $\{m_1, n_1, k_1\}$, denoted by $\{m_2, n_2, k_2\}$, satisfying the fixed-point message equations:*

$$\lim_{m_2,n_2,k_2\to\infty} \mu^{(m_2,n_2,k_2)}_{e\to f_i,\mathrm{SPA}}(x_e) = \lim_{m_2,n_2,k_2\to\infty} \frac{\mu^{(m_2,n_2,k_2)}_{e\to f_i}(x_e)}{Z^{(m_2,n_2,k_2)}_{\mu_{e\to f_i}}} \in \mathbb{R}_{\geq 0}, \qquad x_e \in \mathcal{S}^c_e,\, e = (f_i, f_j) \in \mathcal{E}, \tag{141}$$

$$\lim_{m_2,n_2,k_2\to\infty} \mu^{(m_2,n_2,k_2)}_{e\to f_j,\mathrm{SPA}}(x_e) = \lim_{m_2,n_2,k_2\to\infty} \frac{\mu^{(m_2,n_2,k_2)}_{e\to f_j}(x_e)}{Z^{(m_2,n_2,k_2)}_{\mu_{e\to f_j}}} \in \mathbb{R}_{\geq 0}, \qquad x_e \in \mathcal{S}^c_e,\, e = (f_i, f_j) \in \mathcal{E}, \tag{142}$$

$$\lim_{m_2,n_2,k_2\to\infty} F^{(1)}_{\mathrm{B,p},\mathsf{N}}\big(\boldsymbol{b}^{(m_2,n_2,k_2)}_{\mathcal{F}}\big) = F^*_{\mathrm{B,p},\mathsf{N}} = F^{\mathrm{alt},*}_{\mathrm{B,d},\mathsf{N}} = F^*_{\mathrm{B,d},\mathsf{N}}, \tag{143}$$

$$\Big(\lim_{m_2,n_2,k_2\to\infty} \boldsymbol{b}^{(m_2,n_2,k_2)}_{\mathcal{F}}\Big) \in \mathcal{B}_{\mathcal{F}}(\mathsf{N}). \tag{144}$$

where the vector $\boldsymbol{b}^{(m_2,n_2,k_2)}_f$ is given in (37)*, $\mu^{(m_2,n_2,k_2)}_{e\to f}$ is defined in* (137)*, the constant $Z^{(m_2,n_2,k_2)}_{\mu_{e\to f}}$ is defined in* (138)*, $\mu^{(m_2,n_2,k_2)}_{e\to f,\mathrm{SPA}}$ is defined in* (139)*, the function $F^{(1)}_{\mathrm{B,p},\mathsf{N}}$ is given in* (129)*, the quantity $F^*_{\mathrm{B,p},\mathsf{N}}$ is defined in* (13)*, the quantity $F^{\mathrm{alt},*}_{\mathrm{B,d},\mathsf{N}}$ is defined in* (22)*, the quantity $F^*_{\mathrm{B,d},\mathsf{N}}$ is defined in* (20)*, and the set $\mathcal{B}_{\mathcal{F}}(\mathsf{N})$ is defined in* (9)*. It shows that $\boldsymbol{\mu}^{(m_2,n_2,k_2)} \triangleq \{\boldsymbol{\mu}^{(m_2,n_2,k_2)}_{e\to f}\}_{e\in\partial f, f\in\mathcal{F}}$ converges to SPA fixed-point messages as $m_2, n_2, k_2 \to \infty$.*

Proof. See Appendix S. ■

VIII. Conclusion

In this paper, we have considered a general S-NFG, where the local functions are nonnegative real-valued, and we have related the global minimum of the associated Bethe free energy function to an SPA fixed point. Note that finding the minimum of the Bethe free energy function for an S-NFG is equivalent to finding the associated Bethe partition function. We have developed two main techniques to obtain the main results in this paper.

1) The first technique transforms the Bethe partition function into a maximin optimization problem that can be viewed as the dual formulation of the Bethe partition function.
2) The second technique is used to study relationship between the sequence of the variables, i.e., $\{\boldsymbol{\gamma}^{(m)}, \boldsymbol{\lambda}^{(n)}, \alpha^{(k)}\}_{m,n,k\in\mathbb{Z}_{>0}}$ and the locations of the optimal value for the dual formulation, where the sequences in $\{\boldsymbol{\gamma}^{(m)}, \boldsymbol{\lambda}^{(n)}, \alpha^{(k)}\}_{m,n,k\in\mathbb{Z}_{>0}}$ have been defined to be the sequences that converge to one of the locations of the optimal value of the dual function. Because in the dual formulation, the feasible sets of the collection of variables $\boldsymbol{\lambda}$ and the variable α are the fields of real numbers $\mathbb{R}$, which do not contain $\{\pm\infty\}$. We want to know whether the first-order necessary optimality condition still holds when some of the entries in $\boldsymbol{\lambda}^{(n)}$ and $\alpha^{(k)}$ go to infinity as $n \to \infty$ and $k \to \infty$. We have studied the Taylor series expansion of the objective function in the dual formulation in this case and have shown that the sequences $\{\boldsymbol{\lambda}^{(n)}\}_{n\in\mathbb{Z}_{>0}}$ and $\{\alpha^{(k)}\}_{k\in\mathbb{Z}_{>0}}$ satisfy the generalized first-order necessary optimality condition.

With these two techniques, we have shown the existence of a sequence of messages such that this sequence converges to SPA fixed-point messages, and the associated sequence of beliefs obtained by this sequence of messages converges to one of the locations of the minimum of the Bethe free energy function.

Appendix A
Proof of Proposition 13

The proof of the first statement is given as follows. Because

$$\log\left(\sum_{x_e}\gamma_e(x_e)\right) = 0, \qquad \boldsymbol{\gamma}_e \in \mathcal{B}_e^{\geq},\ \sum_{x_e}\gamma_e(x_e),\ e\in\mathcal{E},$$

we have

$$\begin{aligned} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma},\boldsymbol{\lambda}) &= F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma},\boldsymbol{\lambda}) + \sum_e \log\left(\sum_{x_e}\gamma_e(x_e)\right) \\ &= -\log\left(\frac{\prod_f Z_f(\boldsymbol{\gamma}_{\partial f},\boldsymbol{\lambda}_{\partial f})}{\prod_e\left(\sum_{x_e}\gamma_e(x_e)\right)}\right) \\ &\overset{(a)}{=} -\log\left(\prod_f\left(\sum_{\boldsymbol{x}_f} f(\boldsymbol{x}_f)\cdot\prod_{e\in\partial f}\left(\exp\big(\lambda_{e,f}(x_e)\big)\cdot\sqrt{\frac{\gamma_e(x_e)}{\sum_{x_e}\gamma_e(x_e)}}\right)\right)\right), \end{aligned}$$

where step (a) follows from the definition of Z_f in (16) and the fact that each edge connects two function nodes. With this, the optimization problem (22) is equivalent to the optimization problem

$$\begin{aligned} F^{\mathrm{alt},*}_{\mathrm{B,d,N}} = \inf_{\boldsymbol{\gamma}}\sup_{\boldsymbol{\lambda}}&\left\{F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma},\boldsymbol{\lambda}) + \sum_e\log\left(\sum_{x_e}\gamma_e(x_e)\right)\right\} \\ \text{s.t.}\quad & \lambda_e(x_e)\in\mathbb{R},\ \gamma_e(x_e)\in\mathbb{R}_{\geq 0},\ x_e\in\mathcal{X}_e,\ \sum_{x_e}\gamma_e(x_e)\in\mathbb{R}_{>0},\ e\in\mathcal{E}. \end{aligned}$$

Now we prove the second statement. Because the considered vector $\boldsymbol{\gamma}_e$ is in $\mathcal{B}_e^{>}$, it satisfies $\gamma_e(x_e) > 0$ for all $x_e \in \mathcal{X}_e$. Then the partial derivatives of the objective function in (26) w.r.t. $\boldsymbol{\gamma}$ are given by

$$\frac{\partial}{\partial\gamma_e(x_e)}F^{\mathrm{alt}}_{\mathrm{B,d,N}} + \frac{1}{\sum_{x_e}\gamma_e(x_e)} \overset{(a)}{=} -\frac{b_{f_i,e}(x_e)+b_{f_j,e}(x_e)}{2\gamma_e(x_e)} + 1, \qquad x_e\in\mathcal{X}_e,\ e=(f_i,f_j)\in\mathcal{E},$$

where step (a) follows from the expression of $b_{f,e}$ in (25) for $e \in \partial f$ and $f \in \mathcal{F}$ and the fact that $\boldsymbol{\gamma}_e \in \mathcal{B}_e^{>}$, i.e., $\sum_{x_e} \gamma_e(x_e) = 1$ for all $e \in \mathcal{E}$. Setting the above partial derivatives to zero, which is equivalent to using the condition $\frac{\partial}{\partial \gamma_e(x_e)} F_{\mathrm{B,d,N}}^{\mathrm{alt}} + 1 = 0$ as mentioned in the proposition, we have

$$\gamma_e(x_e) = \frac{b_{f_i,e}(x_e) + b_{f_j,e}(x_e)}{2}, \qquad x_e \in \mathcal{X}_e,\, e = (f_i, f_j) \in \mathcal{E}. \tag{145}$$

Also the partial derivatives of the objective function in (26) w.r.t. $\boldsymbol{\lambda}$ are given by

$$\frac{\partial}{\partial \lambda_e(x_e)} F_{\mathrm{B,d,N}}^{\mathrm{alt}} = -b_{f_i,e}(x_e) + b_{f_j,e}(x_e), \qquad x_e \in \mathcal{X}_e,\, e = (f_i, f_j) \in \mathcal{E}.$$

Setting the above partial derivatives to zero, which is equivalent to using the condition $\frac{\partial}{\partial \lambda_e(x_e)} F_{\mathrm{B,d,N}}^{\mathrm{alt}} = 0$ as mentioned in the proposition, we have

$$b_{f_i,e}(x_e) = b_{f_j,e}(x_e), \qquad x_e \in \mathcal{X}_e,\, e = (f_i, f_j) \in \mathcal{E}. \tag{146}$$

At a stationary point of the objective function, i.e., when both (145) and (146) hold, we have

$$\gamma_e(x_e) = b_{f_i,e}(x_e) = b_{f_j,e}(x_e), \qquad x_e \in \mathcal{X}_e,\, e = (f_i, f_j) \in \mathcal{E}. \tag{147}$$

Because of $\boldsymbol{\lambda} \in \mathbb{R}^{|\mathcal{X}|}$, we have $\exp(\lambda_e(x_e)) \in \mathbb{R}_{>0}$ for all $x_e \in \mathcal{X}_e$ and $e \in \mathcal{E}$ and

$$b_{f_i,e}(x_e) \propto \exp(\lambda_{e,f_i}(x_e)) \cdot \sqrt{\gamma_e(x_e)} \cdot \sum_{\boldsymbol{z}_{f_i} :\, z_e = x_e} f_i(\boldsymbol{z}_{f_i}) \cdot \prod_{e' \in \partial f_i \setminus \{e\}} \Big(\exp\big(\lambda_{e',f_i}(z_{e'})\big) \cdot \sqrt{\gamma_{e'}(z_{e'})} \Big), \tag{148}$$

$$b_{f_j,e}(x_e) \propto \exp\big(\lambda_{e,f_j}(x_e)\big) \cdot \sqrt{\gamma_e(x_e)} \cdot \sum_{\boldsymbol{z}_{f_j} :\, z_e = x_e} f_j(\boldsymbol{z}_{f_j}) \cdot \prod_{e' \in \partial f_j \setminus \{e\}} \Big(\exp\big(\lambda_{e',f_j}(z_{e'})\big) \cdot \sqrt{\gamma_{e'}(z_{e'})} \Big), \tag{149}$$

where λ_{e,f_i} and λ_{e,f_j} are defined in (15). Then we have

$$\begin{aligned} \exp(\lambda_{e,f_i}(x_e)) \cdot \sqrt{\gamma_e(x_e)} &\overset{(a)}{=} \frac{b_{f_j,e}(x_e)}{\exp\big(\lambda_{e,f_j}(x_e)\big) \cdot \sqrt{\gamma_e(x_e)}} \\ &\overset{(b)}{\propto} \sum_{\boldsymbol{z}_{f_j} :\, z_e = x_e} f_j(\boldsymbol{z}_{f_j}) \cdot \prod_{e' \in \partial f_j \setminus \{e\}} \Big(\exp\big(\lambda_{e',f_j}(z_{e'})\big) \cdot \sqrt{\gamma_{e'}(z_{e'})} \Big). \end{aligned}$$

where step (a) follows from the expressions in (147) and the definitions of λ_{e,f_i} and λ_{e,f_j} in (15) and where step (b) follows from (149). Similarly, we have

$$\exp\big(\lambda_{e,f_j}(x_e)\big) \cdot \sqrt{\gamma_e(x_e)} \propto \sum_{\boldsymbol{z}_{f_i} :\, z_e = x_e} f_i(\boldsymbol{z}_{f_i}) \cdot \prod_{e' \in \partial f_i \setminus \{e\}} \Big(\exp\big(\lambda_{e',f_i}(z_{e'})\big) \cdot \sqrt{\gamma_{e'}(z_{e'})} \Big).$$

By letting

$$\mu_{e \to f}(x_e) \propto \exp(\lambda_{e,f}(x_e)) \cdot \sqrt{\gamma_e(x_e)}, \qquad x_e \in \mathcal{X}_e,\, e \in \partial f,\, f \in \mathcal{F},$$

we can obtain a collection of SPA fixed-point messages. The definition of SPA fixed-point messages is given in Definition 5.

APPENDIX B
PROOF OF THEOREM 14

Because of $F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) = F_{\mathrm{B,d,N}}(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ as defined in (23), it is sufficient to prove the properties for $F_{\mathrm{B,d,N}}^{\mathrm{alt}}$ only. The proof for each property for $F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ is listed as follows.

1) [8, Section 4] For fixed $\boldsymbol{\gamma}_e \in \mathcal{B}_e(\mathsf{N})$ for all $e \in \mathcal{E}$, the function $F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ is the negative of a "log-sum-exp" function w.r.t. $\boldsymbol{\lambda}$ and thus it is concave.

2) Although it was proven in [16, Section 6.2], here we provide another approach to prove it. For fixed $\boldsymbol{\lambda}$ and $\{\boldsymbol{\gamma}_{e'}\}_{e' \in \mathcal{E} \setminus \{e\}}$, by the definition of $F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ in (23), it is sufficient to prove that $-\log Z_{f_i}(\boldsymbol{\gamma}_{\partial f_i}, \boldsymbol{\lambda}_{\partial f_i})$ and $-\log Z_{f_j}(\boldsymbol{\gamma}_{\partial f_j}, \boldsymbol{\lambda}_{\partial f_j})$ are convex w.r.t. $\boldsymbol{\gamma}_e$ for $e = (f_i, f_j)$. The function $Z_{f_i}(\boldsymbol{\gamma}_{\partial f_i}, \boldsymbol{\lambda}_{\partial f_i})$ defined in (16) equals

$$Z_{f_i}(\boldsymbol{\gamma}_{\partial f_i}, \boldsymbol{\lambda}_{\partial f_i}) = \sum_{\boldsymbol{x}_{f_i}} f_i(\boldsymbol{x}_{f_i}) \cdot \sqrt{\gamma_e(x_{e,f_i})} \cdot \left(\exp\left(\sum_{e' \in \partial f_i} \lambda_{e',f_i}(x_{e',f_i}) \right) \prod_{e'' \in \partial f_i \setminus \{e\}} \sqrt{\gamma_{e''}(x_{e'',f_i})} \right).$$

Because

- the function $f(\boldsymbol{x}_{f_i})$ is nonnegative for $\boldsymbol{x}_{f_i} \in \mathcal{X}_{f_i}$,
- the function $\sqrt{\gamma_e(x_{e,f_i})}$ is concave w.r.t. $\gamma_e(x_{e,f_i})$,
- the vector $\boldsymbol{\lambda}$ is real-valued,
- the vector $\boldsymbol{\gamma}$ is nonnegative,

the function Z_{f_i} is concave w.r.t. $\boldsymbol{\gamma}_e$. Due to the fact that the logarithm function is a concave and non-decreasing function for positive real arguments, taking the logarithm of a concave function is again a concave function [21, Section 3.2.4]. Then we know that the function $\log Z_{f_i}$ is concave w.r.t. $\boldsymbol{\gamma}_e$ as well. Thus $-\log Z_{f_i}(\boldsymbol{\gamma}_{\partial f_i}, \boldsymbol{\lambda}_{\partial f_i})$ is convex w.r.t. $\boldsymbol{\gamma}_e$. Similarly, we can prove that $-\log Z_{f_j}(\boldsymbol{\gamma}_{\partial f_j}, \boldsymbol{\lambda}_{\partial f_j})$ is also convex w.r.t. $\boldsymbol{\gamma}_e$. Thus the function $F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) = -\sum_f \log Z_f(\boldsymbol{\gamma}_{\partial f}, \boldsymbol{\lambda}_{\partial f})$ is convex w.r.t. $\boldsymbol{\gamma}_e$ for fixed $\boldsymbol{\lambda}$ and $\{\boldsymbol{\gamma}_{e'}\}_{e' \in \mathcal{E} \setminus \{e\}}$.

Appendix C
Proof of Proposition 16

In this section, we prove that the Bethe partition function $Z_{\mathrm{B,p,N}}^{*}$ defined in (12) for an S-NFG N can be written as

$$Z_{\mathrm{B,p,N}}^{*} = \exp\left(-\inf_{\boldsymbol{\gamma}} \sup_{\boldsymbol{\lambda}} \left(-\sum_{f} \log Z_f(\boldsymbol{\gamma}_{\partial f}, \boldsymbol{\lambda}_{\partial f}) \right) \right),$$
$$\text{s.t.} \quad \lambda_e(x_e) \in \mathbb{R}, \qquad x_e \in \mathcal{X}_e,\ \boldsymbol{\gamma}_e \in \mathcal{B}_e^{>},\ e \in \mathcal{E},$$

where Z_f is defined in (16) and $\mathcal{B}_e^{>}$ is defined in (6).

Lemma 62. *For each $\boldsymbol{\beta}_e \in \mathcal{B}_e^{\geq}$, where $\mathcal{B}_e^{\geq}$ is defined in* (5)*, we have*

$$-\sum_{x_e} \beta_e(x_e) \log \beta_e(x_e) = \inf_{\boldsymbol{\gamma}_e} \left(-\sum_{x_e} \beta_e(x_e) \cdot \log \gamma_e(x_e) \right)$$
$$\text{s.t.} \quad \boldsymbol{\gamma}_e \in \mathcal{B}_e^{>}. \tag{150}$$

Proof. It is well-known that (see, e.g., [21, page 222])

$$-\sum_{x_e} \beta_e(x_e) \log \beta_e(x_e) = \inf_{\{\zeta_e(x_e)\}_{x_e}} \left(-\sum_{x_e} \beta_e(x_e) \cdot \zeta_e(x_e) + \log\left(\sum_{x_e} \exp\bigl(\zeta_e(x_e)\bigr) \right) \right) \tag{151}$$
$$\text{s.t.} \quad \zeta_e(x_e) \in \mathbb{R},\ x_e \in \mathcal{X}_e, \tag{152}$$

where the location of the optimal value for the optimization problem defined on the RHS of the above expression is given by

$$\frac{\exp(\zeta_e(x_e))}{\sum_{z_e} \exp\bigl(\zeta_e(z_e)\bigr)} = \beta_e(x_e), \qquad x_e \in \mathcal{X}_e.$$

Thus it is sufficient to consider $\zeta_e(x_e) \in \mathbb{R}$ for all $x_e \in \mathcal{X}_e$ such that $\sum_{x_e} \exp(\zeta_e(x_e)) = 1$ in the above problem. Therefore, the optimization problem (152) becomes

$$-\sum_{x_e} \beta_e(x_e) \log \beta_e(x_e) = \inf_{\{\zeta_e(x_e)\}_{x_e}} -\sum_{x_e} \beta_e(x_e) \cdot \zeta_e(x_e)$$

$$\text{s.t.} \quad \zeta_e(x_e) \in \mathbb{R},\ x_e \in \mathcal{X}_e,\ \sum_{x_e} \exp(\zeta_e(x_e)) = 1.$$

Substituting $\zeta_e(x_e)$ by $\log \gamma_e(x_e)$ for all $x_e \in \mathcal{X}_e$ yields (150). Note that the constraints on $\zeta_e(x_e) \in \mathbb{R}$ for all $x_e \in \mathcal{X}_e$ are equivalent to $\boldsymbol{\gamma}_e \in \mathcal{B}_e^{>}$. ■

By Lemma 62 and the fact that the local consistency constraints in $\mathcal{B}(\mathsf{N})$ are equivalent to

$$\beta_e(x_e) = \frac{1}{2}\Bigg(\beta_{f_i,e}(x_e) + \beta_{f_j,e}(x_e)\Bigg), \qquad x_e \in \mathcal{X}_e,\, e = (f_i, f_j) \in \mathcal{E}, \tag{153}$$

where the marginals $\beta_{f_i,e}(x_e)$ and $\beta_{f_j,e}(x_e)$ are given in (8), we can replace $\{\boldsymbol{\beta}_e\}_e$ with $\boldsymbol{\beta}_{\mathcal{F}}$, both in $F^{*}_{\mathrm{B,p},\mathsf{N}}$ defined in (10) and the constraints (7). Then we transform $Z^{*}_{\mathrm{B,p},\mathsf{N}}$ defined in (12) into a constrained optimization problem w.r.t. the variables $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}_{\mathcal{F}}$, i.e.,

$$Z^{*}_{\mathrm{B,p},\mathsf{N}} = \exp\Bigg(-\min_{\boldsymbol{\beta}_{\mathcal{F}}} \inf_{\boldsymbol{\gamma}} F^{(2)}_{\mathrm{B,p},\mathsf{N}}(\boldsymbol{\gamma}, \boldsymbol{\beta}_{\mathcal{F}})\Bigg) \tag{P1}$$

$$\text{s.t.} \quad \boldsymbol{\gamma}_e \in \mathcal{B}_e^{>}, \quad e \in \mathcal{E}, \tag{154}$$

$$\boldsymbol{\beta}_f \in \mathcal{B}_f, \quad f \in \mathcal{F}, \tag{155}$$

$$\beta_{f_i,e}(x_e) = \beta_{f_j,e}(x_e), \quad x_e \in \mathcal{X}_e,\, i < j,\, e = (f_i, f_j) \in \mathcal{E}, \tag{156}$$

where $\mathcal{B}_f$ is defined in (4), and $F^{(2)}_{\mathrm{B,p},\mathsf{N}}$ is defined to be

$$F^{(2)}_{\mathrm{B,p},\mathsf{N}} : \mathbb{R}^{|\mathcal{X}|}_{>0} \times \prod_f \mathcal{X}_f \to \mathbb{R} \tag{157}$$

$$(\boldsymbol{\gamma}, \boldsymbol{\beta}_{\mathcal{F}}) \mapsto -\sum_f \sum_{\boldsymbol{x}_f} \beta_f(\boldsymbol{x}_f) \cdot \log f(\boldsymbol{x}_f) + \sum_f \sum_{\boldsymbol{x}_f} \beta_f(\boldsymbol{x}_f) \cdot \log \beta_f(\boldsymbol{x}_f)$$

$$- \sum_e \sum_{x_e} \frac{1}{2} \cdot \big(\log \gamma_e(x_e)\big) \cdot \big(\beta_{f_i,e}(x_e) + \beta_{f_j,e}(x_e)\big). \tag{158}$$

The optimization problem (P1) is equivalent to the following optimization problem:

$$Z^{*}_{\mathrm{B,p},\mathsf{N}} = \exp\Bigg(-\inf_{\boldsymbol{\gamma}} \min_{\boldsymbol{\beta}_{\mathcal{F}}} F^{(2)}_{\mathrm{B,p},\mathsf{N}}(\boldsymbol{\gamma}, \boldsymbol{\beta}_{\mathcal{F}})\Bigg) \quad \text{s.t. (154)–(156) hold.} \tag{159}$$

Because

1) the constraints (155) and (156) form a compact set of $\boldsymbol{\beta}_{\mathcal{F}}$;
2) $F^{(2)}_{\mathrm{B,p},\mathsf{N}}$ is continuous w.r.t. $\boldsymbol{\beta}_{\mathcal{F}}$ in this set;
3) $F^{(2)}_{\mathrm{B,p},\mathsf{N}}$ is bounded in this set for fixed $\boldsymbol{\gamma}$,

we have, for fixed $\boldsymbol{\gamma}$,

$$\Big(\min_{\boldsymbol{\beta}_{\mathcal{F}}} F^{(2)}_{\mathrm{B,p},\mathsf{N}}(\boldsymbol{\gamma}, \boldsymbol{\beta}_{\mathcal{F}})\Big) \in \mathbb{R} \quad \text{s.t. (155) and (156) hold, and } \boldsymbol{\gamma}_e \in \mathcal{B}_e^{>}, \quad e \in \mathcal{E}. \tag{160}$$

For fixed $\boldsymbol{\gamma}$, the optimization problem (160) is convex in $\boldsymbol{\beta}_{\mathcal{F}}$, which means that the location of $Z^{*}_{\mathrm{B,p},\mathsf{N}}$ can be obtained using the method of Lagrange multipliers. The details of this method are given as follows. We note that the optimization problem in (159) is equivalent to the following optimization problem:

$$Z^{*}_{\mathrm{B,p},\mathsf{N}} = \exp\Big(\inf_{\boldsymbol{\gamma}} \min_{\boldsymbol{\beta}_{\mathcal{F}}} \sup_{\boldsymbol{\lambda}} \ L(\boldsymbol{\gamma}, \boldsymbol{\beta}_{\mathcal{F}}, \boldsymbol{\lambda})\Big)$$

$$\text{s.t.} \quad \lambda_e(x_e) \in \mathbb{R}, \quad x_e \in \mathcal{X}_e, \quad \boldsymbol{\gamma}_e \in \mathcal{B}_e^{>}, \quad e \in \mathcal{E},$$

$$\boldsymbol{\beta}_f \in \mathcal{B}_f, \quad f \in \mathcal{F},$$

where the Lagrangian function L is given by

$$L(\boldsymbol{\gamma}, \boldsymbol{\beta}_{\mathcal{F}}, \boldsymbol{\lambda}) \triangleq F_{\mathrm{B,p,N}}^{(2)}(\boldsymbol{\gamma}, \boldsymbol{\beta}_{\mathcal{F}}) + \sum_{\substack{e=(f_i,f_j)\\ i<j}} \sum_{x_e \in \mathcal{X}_e} \lambda_e(x_e) \cdot \Big(\beta_{f_j,e}(x_e) - \beta_{f_i,e}(x_e) \Big), \tag{161}$$

and the variables $\boldsymbol{\lambda}$ are the Lagrange multipliers w.r.t. the local consistency constraints in (156). By (160), we know that the optimal value of the following optimization problem is real-valued for fixed $\boldsymbol{\gamma}$:

$$\begin{aligned} &\Big(\min_{\boldsymbol{\beta}_{\mathcal{F}}} \sup_{\boldsymbol{\lambda}} L(\boldsymbol{\beta}_{\mathcal{F}}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) \Big) \in \mathbb{R} \\ &\text{s.t.} \quad \lambda_e(x_e) \in \mathbb{R}, \quad x_e \in \mathcal{X}_e,\, e \in \mathcal{E}, \\ &\boldsymbol{\beta}_f \in \mathcal{B}_f, \quad f \in \mathcal{F}. \end{aligned} \tag{162}$$

Because

1) L is convex in $\boldsymbol{\beta}_{\mathcal{F}}$ for fixed $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$;
2) L is concave w.r.t. $\boldsymbol{\lambda}$ for fixed $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}_{\mathcal{F}}$,

by Sion's minimax theorem (see, e.g., [21]), we can further transform $Z_{\mathrm{B,p,N}}$ into the following optimization problem

$$\begin{aligned} Z_{\mathrm{B,p,N}}^{*} &= \exp\Big(- \inf_{\boldsymbol{\gamma}} \sup_{\boldsymbol{\lambda}} \min_{\boldsymbol{\beta}_{\mathcal{F}}} L(\boldsymbol{\beta}_{\mathcal{F}}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) \Big) \\ &\text{s.t.} \quad \lambda_e(x_e) \in \mathbb{R}, \quad x_e \in \mathcal{X}_e, \quad \boldsymbol{\gamma}_e \in \mathcal{B}_e^{>}, \quad e \in \mathcal{E}, \\ &\boldsymbol{\beta}_f \in \mathcal{B}_f, \quad f \in \mathcal{F}. \end{aligned} \tag{163}$$

Note that in (163), the constraints for $\boldsymbol{\beta}_{\mathcal{F}}$ form a compact set, and L is continuous w.r.t. $\boldsymbol{\beta}_{\mathcal{F}}$. Thus the minimum of L w.r.t. $\{\boldsymbol{\beta}_f\}_{f\in\mathcal{F}}$ for fixed $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ is attainable. By (162) and Sion's minimax theorem, we also know that the optimal value of the following optimization is real-valued for fixed $\boldsymbol{\gamma}$:

$$\begin{aligned} &\Big(\sup_{\boldsymbol{\lambda}} \min_{\boldsymbol{\beta}_{\mathcal{F}}} L(\boldsymbol{\beta}_{\mathcal{F}}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) \Big) \in \mathbb{R} \\ &\text{s.t.} \quad \lambda_e(x_e) \in \mathbb{R}, \quad x_e \in \mathcal{X}_e,\, e \in \mathcal{E}, \\ &\boldsymbol{\beta}_f \in \mathcal{B}_f, \quad f \in \mathcal{F}. \end{aligned} \tag{164}$$

Now we proceed to solve the optimization problem (163) for fixed $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$. Because of the convexity of L w.r.t. $\{\boldsymbol{\beta}_f\}_{f\in\mathcal{F}}$ for fixed $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ and the fact that the constraints for $\boldsymbol{\beta}_{\mathcal{F}}$ form a compact and convex set, the location of the optimal value for the optimization problem in (163), denoted by $\beta_f^*(\boldsymbol{x}_f)$ for all $\boldsymbol{x}_f \in \mathcal{X}_f$ and $f \in \mathcal{F}$, can be obtained by solving the following inequalities [19, Proposition 3.1.1]:

$$\sum_f \big(\nabla_{\boldsymbol{\beta}_f} L \big|_{\boldsymbol{\beta}_f = \boldsymbol{\beta}_f^*} \big)^{\mathsf{T}} \cdot (\boldsymbol{\beta}_f - \boldsymbol{\beta}_f^*) = \sum_f \sum_{\boldsymbol{x}_f} \frac{\partial}{\partial \beta_f(\boldsymbol{x}_f)} L \bigg|_{\boldsymbol{\beta}_f = \boldsymbol{\beta}_f^*} \cdot \big(\beta_f(\boldsymbol{x}_f) - \beta_f^*(\boldsymbol{x}_f) \big) \geq 0, \qquad \forall \boldsymbol{\beta}_f \in \mathcal{B}_f, \quad f \in \mathcal{F}, \tag{165}$$

where $\nabla_{\boldsymbol{\beta}_f} L$ is the gradient of L w.r.t. $\boldsymbol{\beta}_f$, and the entries in the gradient are given by

$$\frac{\partial}{\partial \beta_f(\boldsymbol{x}_f)} L = -\log f(\boldsymbol{x}_f) + \log \beta_f(\boldsymbol{x}_f) + 1 - \sum_{e \in \partial f} \left(\lambda_{e,f}(x_{e,f}) + \frac{1}{2} \log\big(\gamma_e(x_{e,f})\big) \right), \qquad \boldsymbol{x}_f \in \mathcal{X}_f,\, f \in \mathcal{F},$$

where the definition of $\{\boldsymbol{\lambda}_{e,f}\}_{e\in\partial f, f\in\mathcal{F}}$ is given in (15). If we set $\boldsymbol{\beta}_{\mathcal{F}}^*$ to be

$$\beta_f^*(\boldsymbol{x}_f) = \frac{1}{Z_f(\boldsymbol{\gamma}_{\partial f}, \boldsymbol{\lambda}_{\partial f})} \cdot f(\boldsymbol{x}_f) \cdot \prod_{e \in \partial f} \Big(\exp\big(\lambda_{e,f}(x_{e,f})\big) \cdot \sqrt{\gamma_e(x_{e,f})} \Big), \qquad \boldsymbol{x}_f \in \mathcal{X}_f,\, f \in \mathcal{F}, \tag{166}$$

where $Z_f(\boldsymbol{\gamma}_{\partial f}, \boldsymbol{\lambda}_{\partial f})$ is the normalization factor defined in (16), then the resulting $\{\boldsymbol{\beta}_f^*\}_f$ satisfies (166) and is at the location of the optimal value for the optimization problem (163) for fixed $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$. Substituting the above expression (166) for $\{\boldsymbol{\beta}_f\}_{f\in\mathcal{F}}$ into (161), we obtain

$$L(\boldsymbol{\gamma}, \{\boldsymbol{\beta}_f^*\}_f, \boldsymbol{\lambda}) = -\sum_f \log Z_f(\boldsymbol{\gamma}_{\partial f}, \boldsymbol{\lambda}_{\partial f}),$$

which proves (27). Combining with (164), we know that the optimal value of the following optimization problem is real-valued:

$$\left(\sup_{\boldsymbol{\lambda}\in\mathbb{R}^{|\mathcal{X}|}} -\sum_f \log Z_f(\boldsymbol{\gamma}_{\partial f}, \boldsymbol{\lambda}_{\partial f}) \right) \in \mathbb{R}, \qquad \boldsymbol{\gamma} \in \mathbb{R}_{>0}^{|\mathcal{X}|}, \quad \sum_{x_e} \gamma_e(x_e) = 1,\, e\in\mathcal{E},$$

which proves (28).

Appendix D
Proof of Proposition 19

The proof consists of two parts, where the first part analyzes the fixed points of the inner loop, and the second part analyzes the fixed points of the outer loop.

1) By the update rule of $\lambda_{\mathrm{dl},e}^{(t_1)}(x_e)$ in (46), at the fixed point of the inner loop, we have

$$b_{\mathrm{dl},f_i,e}^{(t_1,t_2)}(x_e) = b_{\mathrm{dl},f_j,e}^{(t_1,t_2)}(x_e), \qquad x_e\in\mathcal{X}_e,\, e=(f_i,f_j)\in\mathcal{E}. \tag{167}$$

   By the partial derivatives of $F_{\mathrm{B,d,N_1}}^{\mathrm{alt},*}$ in (40), we know that the fixed point corresponds to a stationary point of $F_{\mathrm{B,d,N_2}}^{\mathrm{alt}}$ with respect to $\boldsymbol{\lambda}$.

2) At the fixed point of the outer loop, we have

$$\frac{1}{2}\cdot\left(b_{\mathrm{dl},f_i,e}^{(t_1,t_2)}(x_e) + b_{\mathrm{dl},f_j,e}^{(t_1,t_2)}(x_e)\right) \overset{(a)}{=} b_{\mathrm{dl},f_i,e}^{(t_1,t_2)}(x_e) = \gamma_{\mathrm{dl},e}^{(t_2)}(x_e) \overset{(a)}{\in} \mathbb{R}_{>0}, \qquad x_e\in\mathcal{X}_e,\, e=(f_i,f_j)\in\mathcal{E}, \tag{168}$$

   where step (a) follows from (167) and where step (b) follows from the fact that $\boldsymbol{\gamma}_e \in \mathcal{B}_e^{>}$ for all $e\in\mathcal{E}$. Then we have

$$\begin{aligned}
\left(\nabla_{\boldsymbol{\gamma}_e} F_{\mathrm{B,d,N_2}}^{\mathrm{alt}}\right)^{\mathsf{T}} \cdot (\boldsymbol{\gamma}_e - \boldsymbol{\gamma}_e')\Big|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_{\mathrm{dl}}^{(t_1)},\, \boldsymbol{\gamma}=\boldsymbol{\gamma}_{\mathrm{dl}}^{(t_2)}} &= \sum_{x_e} \frac{b_{\mathrm{dl},f_i,e}^{(t_1,t_2)}(x_e) + b_{\mathrm{dl},f_j,e}^{(t_1,t_2)}(x_e)}{2\gamma_{\mathrm{dl},e}^{(t_2)}(x_e)} \cdot \left(\gamma_{\mathrm{dl},e}^{(t_2)}(x_e) - \gamma_e'(x_e)\right) \\
&\overset{(a)}{=} \sum_{x_e} \left(\gamma_{\mathrm{dl},e}^{(t_2)}(x_e) - \gamma_e'(x_e)\right) \\
&= 0, \qquad e\in\mathcal{E}, \boldsymbol{\gamma}_e' \in \mathcal{B}_e^{\geq}.
\end{aligned} \tag{169}$$

   where step (a) follows from (168).

Appendix E
Proof of Proposition 21

In Algorithm 2, we can update $\exp\left(\lambda_{\mathrm{dl},e}^{(t_1)}(x_e)\right)\cdot\sqrt{\gamma_{\mathrm{dl},e}^{(t_2)}(x_e)}$ directly instead of updating $\exp\left(\lambda_{\mathrm{dl},e}^{(t_1)}(x_e)\right)$ and $\sqrt{\gamma_{\mathrm{dl},e}^{(t_2)}(x_e)}$, respectively. After updating $\exp\left(\lambda_{\mathrm{dl},e}^{(t_1)}(x_e)\right)\cdot\sqrt{\gamma_{\mathrm{dl},e}^{(t_2)}(x_e)}$ for all $x_e\in\mathcal{E}$, we increase both t_1 and t_2 by one. Note that considering $\exp\left(\lambda_{\mathrm{dl},e}^{(t_1)}(x_e)\right)\cdot\sqrt{\gamma_{\mathrm{dl},e}^{(t_2)}(x_e)}$ is equivalent to considering the message sequence defined in (44) and (45). When we update $\boldsymbol{\mu}^{(t_1,t_2)}$ following the steps in Algorithm 2, the resulting update rules are given as follows.

1) We randomly generate $\mu_{e\to f}^{(0,0)}$ by generating $\gamma_{\mathrm{dl},e}^{(0)}(x_e)$ and $\lambda_{\mathrm{dl},e}^{(0)}(x_e)$ following the uniform distributions in $[0,1)$ and $[-1,1]$, respectively, for all $x_e\in\mathcal{X}_e$ and $e\in\mathcal{E}$.

2) For each $t_1 = t_2 \in \mathbb{Z}_{>0}$ and $e = (f_i, f_j) \in \mathcal{E}$, we update $\mu_{e\to f_i}^{(t_1,t_2)}$ by considering the normalization constraint in (45) and the update rule of $\exp\Bigl(\lambda_{\mathrm{dl},e}^{(t_1)}(x_e)\Bigr) \cdot \sqrt{\gamma_{\mathrm{dl},e}^{(t_2)}(x_e)}$ as stated in (47), (50), and (51):

$$\mu_{e\to f_j}^{(t_1,t_2)}(x_e) \propto \sum_{\boldsymbol{z}_{f_i}:\, z_e = x_e} f_i(\boldsymbol{z}_{f_i}) \cdot \prod_{e' \in \partial f_i \setminus \{e\}} \mu_{e'\to f_i}^{(t_1-1,t_2-1)}(z_{e'}), \tag{170}$$

$$\sum_{e} \mu_{e\to f_j}^{(t_1,t_2)}(x_e) = 1. \tag{171}$$

The update rule of $\mu_{e\to f_j}^{(t_1,t_2)}$ can be obtained similarly.

3) We increase t_1 and t_2 by one, respectively.

4) The update of $\boldsymbol{\mu}^{(t_1,t_2)}(x_e)$ is stopped when some termination criterion is met.

5) Comparing the update rules in (170) and (171) with the update rule in (1) in the SPA in Definition 5, we know that Algorithm 2 is equivalent to the SPA in Definition 5.

6) Similar to the proof of Proposition 19, we can show that each fixed point of Algorithm 2 corresponds to a stationary point of $F_{\mathrm{B,d,N}}^{\mathrm{alt}}$.

APPENDIX F
PROOF OF PROPOSITION 26

We list the proof for each property in the following.

1) Omitted.

2) Omitted.

3) Omitted.

4) Because the S-NFG N_1 is a single-cycle S-NFG, the SPA initialiaztion and update rules specified in Definition 5 for messages $\boldsymbol{\mu}_{1\to f_1}^{(t)}$ and $\boldsymbol{\mu}_{1\to f_2}^{(t)}$ are equivalent to applying the power method for the matrix $\boldsymbol{f}_{1,r} \cdot \boldsymbol{f}_{2,r}^{\mathsf{T}}$. (See, e.g. [22, Section 7.3.1].) Thus at the SPA fixed point, messages $\boldsymbol{\mu}_{1\to f_1}^{(t)}$ and $\boldsymbol{\mu}_{1\to f_2}^{(t)}$ correspond to the eigenvectors associated with the eigenvalue with largest magnitude, i.e., $\Lambda_{\max}(r)$.

5) Based on the previous properties, we can prove this property straightforwardly.

6) By Proposition 16, we have $Z_{\mathrm{B,p,N}_1}^{*} = Z_{\mathrm{B,d,N}_1}^{*}$. Because the S-NFG N_1 is a single-cycle S-NFG and the local functions are all positive-valued, by Proposition 10, the local minima of the Bethe free energy function correspond to the SPA fixed points. Note that the considered S-NFG N_1 has a single SPA fixed point, which means that the collection of beliefs in (66) and (67) evaluated at this fixed point is at the location of the optimal value for the primal formulation. One can verify that

$$-\log(\Lambda_{\max}(r)) = F_{\mathrm{B,p,N}_1}(\boldsymbol{\beta}) \qquad \text{s.t. } \boldsymbol{\beta} \in \mathcal{B}(\mathsf{N}_1), \text{ where } \boldsymbol{\beta}_{f_1} \text{ satisfies (66) and } \boldsymbol{\beta}_{f_2} \text{ satisfies (67).}$$

APPENDIX G
PROOF OF PROPOSITION 27

The proof for each property is listed as follows.

1) The eigenvalue $\Lambda_{\max}(r)$ equals

$$\Lambda_{\max}(r) = 1 + \frac{\delta_2(r)}{2} + \frac{c_1(r)}{2} + \frac{\delta_1(r)\cdot\delta_3(r)}{2}. \tag{172}$$

where $\delta_i(r)$ is defined in (61) and (62) for $i \in [3]$, and $c_1(r)$ is defined in (65). The associated eigenvectors equal

$$\boldsymbol{v}_{\mathrm{L}} \propto \begin{pmatrix} \frac{\delta_2(r)-\delta_1(r)\cdot\delta_3(r)+c_1(r)}{2\left(\delta_3(r)+1\right)} & 1 \end{pmatrix}^{\mathsf{T}} \overset{(a)}{\in} \mathbb{R}_{>0}^{|\mathcal{X}_e|}, \quad \boldsymbol{v}_{\mathrm{R}} \propto \begin{pmatrix} \frac{\delta_2(r)-\delta_1(r)\cdot\delta_3(r)+c_1(r)}{2\left(\delta_1(r)+\delta_2(r)\right)} & 1 \end{pmatrix}^{\mathsf{T}} \overset{(a)}{\in} \mathbb{R}_{>0}^{|\mathcal{X}_e|}, \tag{173}$$

where step (a) follows from the property of $c_1(r)$ in (65). Taking the limit $r \downarrow 0$ completes the proof.

2) By Proposition 26, the SPA fixed-point messages are given by some eigenvectors:

$$\boldsymbol{\mu}_{1\to f_1}^{(t)} \propto \boldsymbol{v}_{\mathrm{L}} \in \mathbb{R}_{>0}^{|\mathcal{X}_e|}, \tag{174}$$

$$\boldsymbol{\mu}_{1\to f_2}^{(t)} \propto \boldsymbol{v}_{\mathrm{R}} \in \mathbb{R}_{>0}^{|\mathcal{X}_e|}, \tag{175}$$

$$\boldsymbol{\mu}_{2\to f_1}^{(t)} \propto \begin{pmatrix} \frac{\delta_1(r)\cdot\delta_3(r)-\delta_2(r)+c_1(r)}{2\left(\delta_1(r)+\delta_2(r)\right)} & 1 \end{pmatrix}^{\mathsf{T}} \overset{(a)}{\in} \mathbb{R}_{>0}^{|\mathcal{X}_e|}, \tag{176}$$

$$\boldsymbol{\mu}_{2\to f_2}^{(t)} \propto \begin{pmatrix} \frac{\delta_1(r)\cdot\delta_3(r)-\delta_2(r)+c_1(r)}{2\left(\delta_3(r)+1\right)} & 1 \end{pmatrix}^{\mathsf{T}} \overset{(a)}{\in} \mathbb{R}_{>0}^{|\mathcal{X}_e|}, \tag{177}$$

where step (a) follows from the property of $c_1(r)$ in (65). Taking the limit $r \downarrow 0$ completes the proof.

3) As stated in Proposition 26, the location of $F_{\mathrm{B,p,N_1}}^{*}$ is given by the SPA fixed point and satisfies (66). By the SPA fixed-point messages given in (174)–(177), there exists a scalar $c_{f_1} \in \mathbb{C}$ s.t.

$$\beta_{f_1}(1,1) = c_{f_1} \cdot \frac{2\delta_1(r) + 2\delta_2(r) + \delta_2(r)\cdot c_1(r) + \left(\delta_2(r)\right)^2 - \delta_1(r)\cdot\delta_2(r)\cdot\delta_3(r)}{2\left(\delta_1(r)+\delta_2(r)\right)},$$

$$\beta_{f_1}(1,2) = c_{f_1} \cdot \frac{2\delta_1(r) + 2\delta_2(r) + \delta_2(r)\cdot c_1(r) + \left(\delta_2(r)\right)^2 - \delta_1(r)\cdot\delta_2(r)\cdot\delta_3(r)}{2\left(\delta_1(r)+\delta_2(r)\right)} \cdot \frac{\delta_2(r) - \delta_1(r)\cdot\delta_3(r) + c_1(r)}{2(\delta_3(r)+1)},$$

$$\beta_{f_1}(2,1) = c_{f_1} \cdot \left(\frac{\delta_1(r)\cdot\left(\delta_1(r)\cdot\delta_3(r) + c_1(r)\right) - \left(\delta_2(r)\right)^2\cdot(2\delta_3(r)+1)}{2\left(\delta_1(r)+\delta_2(r)\right)} + \frac{\delta_1(r)\cdot(2\delta_3(r)+1)}{2} \right),$$

$$\beta_{f_1}(2,2) = \beta_{f_1}(1,1),$$

which proves (69). The proof of (70) is similar and thus it is omitted here.

Appendix H
Proof of Proposition 28

The proof for each property is listed as follows.

1) The function Z_{f_1} defined in (16) can be written as

$$\begin{aligned} Z_{f_1}(\boldsymbol{\gamma},\boldsymbol{\lambda}) &= \exp\left(\lambda_1(1)+\lambda_2(1)\right)\cdot\sqrt{\gamma_1(1)\cdot\gamma_2(1)} \\ &\quad + \exp\left(\lambda_1(1)+\lambda_2(2)\right)\cdot\sqrt{\gamma_1(1)\cdot\gamma_2(2)} \\ &\quad + \delta_1(r)\cdot\exp\left(\lambda_1(2)+\lambda_2(1)\right)\cdot\sqrt{\gamma_1(2)\cdot\gamma_2(1)} \\ &\quad + \exp\left(\lambda_1(2)+\lambda_2(2)\right)\cdot\sqrt{\gamma_1(2)\cdot\gamma_2(2)}, \end{aligned} \tag{178}$$

where step (a) follows from the definition of $\boldsymbol{\lambda}_{e,f}$ in (15). The function Z_{f_2} can be written as

$$\begin{aligned} Z_{f_2}(\boldsymbol{\gamma},\boldsymbol{\lambda}) &= \exp\left(-\left(\lambda_1(1)+\lambda_2(1)\right)\right)\cdot\sqrt{\gamma_1(1)\cdot\gamma_2(1)} \\ &\quad + \delta_2(r)\cdot\exp\left(-\left(\lambda_1(1)+\lambda_2(2)\right)\right)\cdot\sqrt{\gamma_1(1)\cdot\gamma_2(2)} \\ &\quad + \delta_3(r)\cdot\exp\left(-\left(\lambda_1(2)+\lambda_2(1)\right)\right)\cdot\sqrt{\gamma_1(2)\cdot\gamma_2(1)} \\ &\quad + \exp\left(-\left(\lambda_1(2)+\lambda_2(2)\right)\right)\cdot\sqrt{\gamma_1(2)\cdot\gamma_2(2)}. \end{aligned} \tag{179}$$

If we consider $(\boldsymbol{\gamma},\boldsymbol{\lambda})$ given in (71)–(72), then we have

$$Z_{f_2}(\boldsymbol{\gamma},\boldsymbol{\lambda})\cdot Z_{f_1}(\boldsymbol{\gamma},\boldsymbol{\lambda}) = \Lambda_{\max} \overset{(a)}{=} Z_{\mathrm{B,d,N_1}}^{*},$$

where step (a) follows from Proposition 26. By the SPA fixed-point messages $\boldsymbol{\mu}$ given in (174)–(177) and the property of $\delta_i(r)$ in (61) and (62) for $i \in [3]$, we can prove (73)–(75).

2) We first prove that as for $\boldsymbol{\gamma}$, the location for $Z_{\mathrm{B,d,N}_1}^{\mathrm{alt},*}$ satisfies $\boldsymbol{\gamma}_e \in \mathcal{B}_e^{>}$ for all $e \in \mathcal{E}$. There are various cases that need to be discussed.

   a) We first suppose that $\gamma_1(1) = 0$, which implies $\gamma_1(2) = 1$ by $\boldsymbol{\gamma}_1 \in \mathcal{B}_1^{\geq}$. If we set

$$\exp(\lambda_1(2)) = \exp(\lambda_2(2)) = 1, \qquad \exp(\lambda_2(1)) = \sqrt{\frac{\delta_3(r)}{\delta_1(r)}},$$

   then we have

$$\begin{aligned} Z_{f_1}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) &= Z_{f_2}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) \\ &= \sqrt{\delta_1(r) \cdot \delta_3(r) \cdot \gamma_2(1)} + \sqrt{\gamma_2(2)} \\ &\overset{(a)}{\leq} \sqrt{\delta_1(r) \cdot \delta_3(r) + 1}, \end{aligned}$$

   where step (a) follows from the Cauchy-Schwarz inequality and the fact that $\boldsymbol{\gamma}_2 \in \mathcal{B}_2^{\geq}$. Based on the above derivations, we have

$$\inf_{\boldsymbol{\lambda} \in \mathbb{R}^{|\mathcal{X}|}} Z_{\mathrm{B,d,N}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) = \inf_{\boldsymbol{\lambda} \in \mathbb{R}^{|\mathcal{X}|}} Z_{f_1}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) \cdot Z_{f_2}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) \leq \delta_1(r) \cdot \delta_3(r) + 1 \overset{(a)}{<} \Lambda_{\max} \overset{(b)}{=} Z_{\mathrm{B,d,N}_1}^{*}.$$

   where step (a) follows from the expression of $\Lambda_{\max}$ in (172) and the property of $c_1(r)$ in (65) and where step (b) follows from (68).

   b) The proofs for other cases where at least one of the entries in $\boldsymbol{\gamma}$ equals zero, are similar and thus they are omitted here. Consequently, we know that if at least one of the entries in $\boldsymbol{\gamma}$ equals zero, we have

$$\inf_{\boldsymbol{\lambda} \in \mathbb{R}^{|\mathcal{X}|}} Z_{\mathrm{B,d,N}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) < \Lambda_{\max} = Z_{\mathrm{B,d,N}_1}^{*}.$$

   Because of the above derivations and

$$\Lambda_{\max} = Z_{\mathrm{B,d,N}_1}^{*} \overset{(a)}{=} \sup_{\boldsymbol{\gamma} \in \mathcal{B}_1^{\geq} \times \mathcal{B}_2^{\geq}} \inf_{\boldsymbol{\lambda} \in \mathbb{R}^{|\mathcal{X}|}} Z_{\mathrm{B,d,N}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}),$$

   where step (a) follows from the definition of $Z_{\mathrm{B,d,N}_1}^{*}$ in (19), we have

$$\sup_{\boldsymbol{\gamma} \in \mathcal{B}_1^{\geq} \times \mathcal{B}_2^{\geq}} \inf_{\boldsymbol{\lambda} \in \mathbb{R}^{|\mathcal{X}|}} Z_{\mathrm{B,d,N}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) = \Lambda_{\max} = Z_{\mathrm{B,d,N}_1}^{*}.$$

   By Proposition 15, we know that for the considered S-NFG N_1, we have $Z_{\mathrm{B,d,N}_1}^{\mathrm{alt},*} = Z_{\mathrm{B,d,N}_1}^{*} = \Lambda_{\max}$.

APPENDIX I
PROOF OF PROPOSITION 35

We prove each property separately.

1) Recall that we consider $\boldsymbol{\gamma}_e \in \mathcal{B}_e^{\geq}$ for all $e \in \mathcal{E}$. By the expressions of Z_{f_1} in (83) and Z_{f_2} in (84), we have

$$\begin{aligned} Z_{f_1}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) \cdot Z_{f_2}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) &\overset{(a)}{\geq} \gamma_1(1) \cdot \gamma_2(1) + \gamma_1(2) \cdot \gamma_2(2) \\ &\quad + \left(\exp\left(\sum_e \big(\lambda_e(1) - \lambda_e(2)\big) \right) + \exp\left(- \sum_e \big(\lambda_e(1) - \lambda_e(2)\big) \right) \right) \cdot \sqrt{\gamma_1(1) \cdot \gamma_1(2) \cdot \gamma_2(1) \cdot \gamma_2(2)} \\ &\overset{(b)}{\geq} \left(\sqrt{\gamma_1(1) \cdot \gamma_2(1)} + \sqrt{\gamma_1(2) \cdot \gamma_2(2)} \right)^2, \end{aligned} \tag{180}$$

where step (a) follows from the fact that the terms in $Z_{f_1}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) \cdot Z_{f_2}(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ are nonnegative, where step (b) follows from the fact that $x + 1/x \geq 2$ for $x \in \mathbb{R}_{>0}$, i.e.,

$$\exp\left(\sum_e \bigl(\lambda_e(1) - \lambda_e(2)\bigr)\right) + \exp\left(-\sum_e \bigl(\lambda_e(1) - \lambda_e(2)\bigr)\right) \geq 2.$$

By the expression of $F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ in (54) and the inequality in (180), we have

$$F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) \leq -2\log\Bigl(\sqrt{\gamma_1(1)\cdot\gamma_2(1)} + \sqrt{\gamma_1(2)\cdot\gamma_2(2)}\Bigr), \qquad \boldsymbol{\lambda} \in \mathbb{R}^{|\mathcal{X}|}. \tag{181}$$

Also by the definition of $\boldsymbol{\lambda}^{(n)}$ in Definition 33, we have

$$\lim_{n\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}^{(n)}) = \lim_{n\to\infty} -\log\Bigl(Z_{f_1}(\boldsymbol{\gamma}, \boldsymbol{\lambda}^{(n)}) \cdot Z_{f_2}(\boldsymbol{\gamma}, \boldsymbol{\lambda}^{(n)})\Bigr) = -2\log\Bigl(\sqrt{\gamma_1(1)\cdot\gamma_2(1)} + \sqrt{\gamma_1(2)\cdot\gamma_2(2)}\Bigr). \tag{182}$$

Then we can prove the first property:

$$\begin{aligned}
-2\log\Bigl(\sqrt{\gamma_1(1)\cdot\gamma_2(1)} + \sqrt{\gamma_1(2)\cdot\gamma_2(2)}\Bigr) &\overset{(a)}{=} \lim_{n\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}^{(n)}) \\
&\leq \sup_{\boldsymbol{\lambda}\in\mathbb{R}^{|\mathcal{X}|}} F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) \\
&\overset{(b)}{=} F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma}) \\
&\overset{(c)}{\leq} -2\log\Bigl(\sqrt{\gamma_1(1)\cdot\gamma_2(1)} + \sqrt{\gamma_1(2)\cdot\gamma_2(2)}\Bigr),
\end{aligned}$$

where step (a) follows from the expressions in (182), where step (b) follows from the definition of $F^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma})$ in (29) for N_1, and where step (c) follows from the inequality in (181).

2) By the expression of $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma})$ in (86), for all $\boldsymbol{\gamma} \in \prod_e \mathcal{B}_e^{\geq}$, we have

$$\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma}) = -2\log\Bigl(\sqrt{\gamma_1(1)\cdot\gamma_2(1)} + \sqrt{\gamma_1(2)\cdot\gamma_2(2)}\Bigr) \overset{(a)}{\geq} -2\log\Bigl(\gamma_1(1)+\gamma_1(2)\Bigr) - 2\log\Bigl(\gamma_2(1)+\gamma_2(2)\Bigr) \overset{(b)}{=} 0, \tag{183}$$

where step (a) follows from the Cauchy-Schwarz inequality and where step (b) follows from the fact that $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathcal{B}_1^{\geq}$ as defined in (14), i.e.,

$$\gamma_1(1) + \gamma_1(2) = 1,\ \gamma_2(1) + \gamma_2(2) = 1.$$

Then by the definition of $F^{\mathrm{alt},*}_{\mathrm{B,d,N_1}}$ in (22), we get $F^{\mathrm{alt},*}_{\mathrm{B,d,N_1}} = 0$. Then it holds that

$$F^{\mathrm{alt},*}_{\mathrm{B,d,N_1}} \overset{(a)}{=} F^{*}_{\mathrm{B,p,N}} \overset{(b)}{=} F^{*}_{\mathrm{B,d,N_1}} \overset{(c)}{=} \hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma}^{(1)}) = 0.$$

where step (a) follows from (82), where step (b) follows from Proposition 16, and where step (c) follows from the definition of $\boldsymbol{\gamma}^{(1)}$ in Definition 33 and the expression of $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N_1}}(\boldsymbol{\gamma}^{(1)})$ in (183).

Appendix J
Proof of Lemma 36

We only prove for function node f_1. The proof for f_2 is similar and thus it is omitted here.

By the expression of $\boldsymbol{b}_{f_1}^{(1,n)}$ in (57) and the definition of $\boldsymbol{\gamma}^{(1)}$ and $\boldsymbol{\lambda}^{(n)}$ in Definition 33, we have

$$\boldsymbol{b}_{f_1}^{(1,n)} = \frac{1}{\gamma_1^{(1)}(1) + \exp(-n)\cdot\sqrt{\gamma_1^{(1)}(1)\cdot\gamma_1^{(1)}(2)} + \gamma_1^{(1)}(2)} \begin{pmatrix} \gamma_1^{(1)}(1) & \exp(-n)\cdot\sqrt{\gamma_1^{(1)}(1)\cdot\gamma_1^{(1)}(2)} \\ 0 & \gamma_1^{(1)}(2) \end{pmatrix}.$$

Taking the limit $n \to \infty$, we have

$$\lim_{n\to\infty} \boldsymbol{b}_{f_1}^{(1,n)} = \frac{1}{\gamma_1^{(1)}(1) + \gamma_1^{(1)}(2)} \begin{pmatrix} \gamma_1^{(1)}(1) & 0 \\ 0 & \gamma_1^{(1)}(2) \end{pmatrix} \overset{(a)}{=} \begin{pmatrix} \gamma_1^{(1)}(1) & 0 \\ 0 & \gamma_1^{(1)}(2) \end{pmatrix},$$

where step (a) follows from the property $\boldsymbol{\gamma}_1 \in \mathcal{B}_1^{>}$ as stated in (53). Because

$$\lim_{n\to\infty} b_{f_1,e}^{(1,n)}(x_e) = \lim_{n\to\infty} b_{f_2,e}^{(1,n)}(x_e) = \gamma_e^{(1)}(x_e), \qquad x_e \in \mathcal{X}_e, e \in [2], \tag{184}$$

the beliefs $\boldsymbol{b}_{f_1}^{(1,n)}$ and $\boldsymbol{b}_{f_2}^{(1,n)}$ satisfy the local consistency constraints and so (88) follows.

APPENDIX K
PROOF OF LEMMA 37

By the definition of $\boldsymbol{\gamma}^{(1)}$ in Definition 33 and the definition of $\mathcal{B}_e^{>}$ in (6), we have $\boldsymbol{\gamma}_1^{(1)} \in \mathbb{R}_{>0}^{|\mathcal{X}_1|}$. Following the definition of $\boldsymbol{\lambda}^{(n)}$ given in Definition 33, we have

$$Z_{f_1}\big(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\lambda}^{(n)}\big) \geq \lim_{n\to\infty} Z_{f_1}\big(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\lambda}^{(n)}\big) = \gamma_1^{(1)}(1) + \gamma_1^{(1)}(2) \in \mathbb{R}_{>0}. \tag{185}$$

By the definition of $b_{f_1}^{(1,n)}$ in (55), we have

$$\begin{aligned}
\lim_{n\to\infty} \sum_{x_1,x_2} b_{f_1}^{(1,n)}(x_1,x_2)\cdot(\lambda_1^{(n)}(x_1) + \lambda_2^{(n)}(x_2)) &\overset{(a)}{=} \lim_{n\to\infty} b_{f_1}^{(1,n)}(1,2)\cdot \lambda_1^{(n)}(1) \\
&= \lim_{n\to\infty} \frac{\gamma_1^{(1)}(1)}{Z_{f_1}\big(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\lambda}^{(n)}\big)} \exp(-n)\cdot(-n) \\
&\overset{(b)}{=} 0,
\end{aligned}$$

where step (a) follows from the definition of $\boldsymbol{\lambda}^{(n)}$ in Definition 33, and where at step (b) follows from the fact that $\lim_{n\to\infty}\exp(-n)\cdot(-n) = 0$, and (185), i.e., a product of a sequence that converges to zero and a bounded sequence is again a sequence that converges to zero. Similarly, we obtain

$$\lim_{n\to\infty} \sum_{x_1,x_2} b_{f_2}^{(1,n)}(x_1,x_2)\cdot(\lambda_1^{(n)}(x_1) + \lambda_2^{(n)}(x_2)) = 0.$$

APPENDIX L
PROOF OF THEOREM 38

By Lemma 36, the continuity of the Kullback–Leibler (K-L) divergence function, and the definition of the marginals $b_{f_1,e}$ and $b_{f_2,e}$ in (58), we have

$$\lim_{n\to\infty} \sum_{e\in[2]} \sum_{x_e:\, \gamma_e^{(1)}(x_e)>0} \frac{b_{f_1,e}^{(1,n)}(x_e) + b_{f_2,e}^{(1,n)}(x_e)}{2} \cdot \log\left(\frac{2\gamma_e^{(1)}(x_e)}{b_{f_1,e}^{(1,n)}(x_e) + b_{f_2,e}^{(1,n)}(x_e)} \right) = 0, \qquad m \in \mathbb{Z}_{>0}. \tag{186}$$

By the fact that the limit of a finite sum of sequences equals the finite sum of the limits of the sequences, provided that the limit of each sequence exists in $\mathbb{R}$, we have

$$\begin{aligned}
\lim_{n\to\infty} F_{\mathrm{B,p,N_1}}^{(1)}(\boldsymbol{b}^{(1,n)}) &= \lim_{n\to\infty} -\log\Big(Z_{f_1}\big(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\lambda}^{(n)}\big)\cdot Z_{f_2}\big(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\lambda}^{(n)}\big)\Big) \\
&\quad + \lim_{n\to\infty}\Big(\sum_{x_1,x_2} b_{f_1}^{(1,n)}(x_1,x_2)\cdot(\lambda_1^{(n)}(x_1)+\lambda_2^{(n)}(x_2)) - \sum_{x_1,x_2} b_{f_2}^{(1,n)}(x_1,x_2)\cdot(\lambda_1^{(n)}(x_1)+\lambda_2^{(n)}(x_2))\Big) \\
&\quad + \lim_{n\to\infty} \sum_{e} \sum_{x_e:\, \gamma_1^{(1)}(x_e)>0} \frac{b_{f_1,e}^{(1,n)}(x_e) + b_{f_2,e}^{(1,n)}(x_e)}{2} \cdot \log\left(\frac{2\gamma_e^{(1)}(x_e)}{b_{f_1,e}^{(1,n)}(x_e) + b_{f_2,e}^{(1,n)}(x_e)} \right) \\
&\overset{(a)}{=} 0,
\end{aligned}$$

where step (a) follows from the property in (87), i.e.,

$$\lim_{n\to\infty} -\log\Big(Z_{f_1}\big(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\lambda}^{(n)}\big)\cdot Z_{f_2}\big(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\lambda}^{(n)}\big)\Big) = F_{\mathrm{B,d,N_1}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(1)}) = 0,$$

and Lemma 37, and the equality in (186). By (87), we have

$$\lim_{n\to\infty} F^{(1)}_{\mathrm{B,p,N_1}}(\boldsymbol{b}^{(1,n)}) = F^{*}_{\mathrm{B,p,N_1}} = F^{*}_{\mathrm{B,d,N_1}} = F^{\mathrm{alt},*}_{\mathrm{B,d,N_1}} = 0.$$

Appendix M

Proof of Lemma 46

We consider the following $\boldsymbol{\lambda}$:

$$\exp\bigl(\lambda_e(x_e)\bigr) = \Bigl(W(0|x_e)\Bigr)^{-1/3}, \qquad x_e \in \mathcal{X}_e = \{0,1\},\, e \in \mathcal{E}. \tag{187}$$

Then the functions in (106) and (107) become

$$Z_{f_i}(\boldsymbol{\gamma}_{\partial f_i}, \boldsymbol{\lambda}_{\partial f_i}) = \sum_{\boldsymbol{x}_{f_i}} [\text{all } x_e,\, e\in\partial f_i \text{ are equal}] \cdot \prod_{e\in\partial f_i} \sqrt{\gamma_e(\boldsymbol{x}_{f_i})}, \qquad i\in\mathcal{I},\, \boldsymbol{\gamma}_e \in \mathcal{B}_e^{\geq},\, e\in\partial f_i,$$

$$Z_{f_j}(\boldsymbol{\gamma}_{\partial f_j}, \boldsymbol{\lambda}_{\partial f_j}) = \sum_{\boldsymbol{x}_{f_j}} \left[\sum_{e\in\partial f} x_e = 0 \mod 2\right] \cdot \prod_{e\in\partial f_j} \left(\Bigl(W(0|x_e)\Bigr)^{1/3} \cdot \sqrt{\gamma_e(x_e)}\right), \quad j\in\mathcal{J},\, \boldsymbol{\gamma}_e\in\mathcal{B}_e^{\geq},\, e\in\partial f_j.$$

Then we can find the upper bounds of the above local functions. The proof for each $i \in \mathcal{I}$ and each $j \in \mathcal{J}$ is similar. Thus we only provide the proof for f_1 and f_5.

1) For f_1, because of $0 \leq \gamma_i(x_1) \leq 1$ for all $x_1 \in \mathcal{X}_1$ and $i \in \{1,2,3\}$, we have

$$\begin{aligned} Z_{f_1}(\boldsymbol{\gamma}_{\partial f_1}, \boldsymbol{\lambda}_{\partial f_1}) &= \sum_{x_1\in\{0,1\}} \sqrt{\gamma_1(x_1)\cdot\gamma_2(x_1)\cdot\gamma_3(x_1)} \\ &\leq \sqrt{\gamma_1(0)\cdot\gamma_2(0)} + \sqrt{\gamma_1(1)\cdot\gamma_2(1)} \\ &\leq \sqrt{\gamma_1(0)+\gamma_2(0)} \cdot \sqrt{\gamma_1(1)+\gamma_2(1)} \\ &\overset{(a)}{=} 1\cdot 1, \qquad \boldsymbol{\gamma}_e \in \mathcal{B}_e^{\geq},\, e\in\partial f_1, \end{aligned} \tag{188}$$

where step (a) follows from $\gamma_i(0)+\gamma_i(1) = 1$ for $i\in\{1,2\}$. The proof for other local functions in $\{f_i\}_{i\in\mathcal{I}}$ is similar. Based on that, we have

$$Z_{f_i}(\boldsymbol{\gamma}_{\partial f_i}, \boldsymbol{\lambda}_{\partial f_i}) \leq 1, \qquad i\in\mathcal{I}. \tag{189}$$

2) For f_5, we have

$$Z_{f_5}(\boldsymbol{\gamma}_{\partial f_5}, \boldsymbol{\lambda}_{\partial f_5}) = \Bigl(W(0|0)\Bigr)^{4/3} \prod_{e\in\partial f_5} \sqrt{\gamma_e(0)} + \sum_{\boldsymbol{x}_{f_5}\in\mathcal{X}_{f_5}\setminus\{(0,\ldots,0)\}} \prod_{e\in\partial f_5} \left(\Bigl(W(0|x_e)\Bigr)^{1/3} \cdot \sqrt{\gamma_e(x_e)}\right). \tag{190}$$

where $\mathcal{X}_{f_5}$ defined in (97) is given by

$$\mathcal{X}_{f_5} = \{\boldsymbol{x}_{f_5} \mid x_1+x_4+x_7+x_{10} = 0 \mod 2\}.$$

In particular, we define $\{x_e^*\}_{e\in\partial f_5}$ to be

$$(x_e^*)_{e\in\partial f_5} \in \mathrm{argmax}_{\boldsymbol{x}_{f_5}\in\mathcal{X}_{f_5}\setminus\{(0,\ldots,0)\}} \prod_{e\in\partial f_5} \sqrt{\gamma_e(x_e)}, \qquad \boldsymbol{\gamma}_e \in \mathcal{B}_e^{\geq},\, e\in\partial f_5, \tag{191}$$

Clearly, the Hamming weight of $(x_e^*)_{e\in\partial f_5}$ is larger than or equal to 2. It holds that

$$\begin{aligned} &\sum_{\boldsymbol{x}_{f_5}\in\mathcal{X}_{f_5}\setminus\{(0,\ldots,0)\}} \prod_{e\in\partial f_5} \left(\Bigl(W(0|x_e)\Bigr)^{1/3} \cdot \sqrt{\gamma_e(x_e)}\right) \\ &\leq \left(\sum_{\boldsymbol{x}_{f_5}\in\mathcal{X}_{f_5}\setminus\{(0,\ldots,0)\}} \prod_{e\in\partial f_5} \Bigl(W(0|x_e)\Bigr)^{1/3}\right) \cdot \max_{\boldsymbol{x}_{f_5}\in\mathcal{X}_{f_5}\setminus\{(0,\ldots,0)\}} \prod_{e\in\partial f_5} \sqrt{\gamma_e(x_e)} \end{aligned}$$

$$\overset{(a)}{=} \left(\sum_{\boldsymbol{x}_{f_5} \in \mathcal{X}_{f_5} \setminus \{(0,\ldots,0)\}} \prod_{e \in \partial f_5} \Big(W(0|x_e) \Big)^{1/3} \right) \cdot \prod_{e \in \partial f_5} \sqrt{\gamma_e(x_e^*)}$$
$$\overset{(b)}{\leq} \big(W(0|0) \big)^{4/3} \cdot \prod_{e \in \partial f_5} \sqrt{\gamma_e(x_e^*)}, \qquad \boldsymbol{\gamma}_e \in \mathcal{B}_e^{\geq},\, e \in \partial f_5,$$

where step (a) follows from (191) and where step (b) follows from the assumption about the channel law in (100). Note that $(x_e^*)_{e \in \partial f_5}$ as defined in (191) contains at least two components, denoted by $x_{e_1}^*$ and $x_{e_2}^*$, such that $x_{e_1}^* \neq 0$ and $x_{e_2}^* \neq 0$. Then the expression in (190) satisfies

$$Z_{f_5}(\boldsymbol{\gamma}_{\partial f_5}, \boldsymbol{\lambda}_{\partial f_5}) \leq \Big(W(0|0) \Big)^{4/3} \cdot \left(\prod_{e \in \partial f_5} \sqrt{\gamma_e(0)} + \prod_{e \in \partial f_5} \sqrt{\gamma_e(x_e^*)} \right)$$
$$\overset{(a)}{\leq} \Big(W(0|0) \Big)^{4/3} \cdot \left(\sqrt{\gamma_{e_1}(0) \cdot \gamma_{e_2}(0)} + \sqrt{\gamma_{e_1}(x_{e_1}^*) \cdot \gamma_{e_2}(x_{e_2}^*)} \right)$$
$$\overset{(b)}{\leq} \Big(W(0|0) \Big)^{4/3}, \qquad \boldsymbol{\gamma}_e \in \mathcal{B}_e^{\geq},\, e \in \partial f_5,$$

where step (a) follows from $\boldsymbol{\gamma}_e \in \mathcal{B}_e^{\geq}$ for all $e \in \partial f_5$:

$$\gamma_e(x_e) \leq 1, \qquad x_e \in \{0,1\},\, e \in \partial f_5,$$

where step (b) follows from the similar considerations as in (188) and thus the details are omitted here. The proof for other elements in $\mathcal{J}$ is similar. Therefore, we have

$$Z_{f_j}(\boldsymbol{\gamma}_{\partial f_j}, \boldsymbol{\lambda}_{\partial f_j}) \leq \Big(W(0|0) \Big)^{4/3}, \qquad j \in \mathcal{J}. \tag{192}$$

When we consider $\boldsymbol{\lambda}$ in (187), the function $Z_{\mathrm{B,d,N_4}}(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ given in (105) satisfies

$$Z_{\mathrm{B,d,N_4}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) \overset{(a)}{\leq} \Big(W(0|0) \Big)^{4},$$

where step (a) follows from the inequalities in (189) and (192). By the expression of $Z_{\mathrm{B,d,N_4}}^{\mathrm{alt},*}$ in (104), we have

$$Z_{\mathrm{B,d,N_4}}^{\mathrm{alt},*} \leq \Big(W(0|0) \Big)^{4}.$$

Appendix N
Proof of Proposition 51

If we set $\boldsymbol{\gamma}$ to be

$$\gamma_e(x_e) \triangleq \begin{cases} 1 & \text{if } x_e = x_e' \\ 0 & \text{otherwise} \end{cases}, \qquad x_e \in \mathcal{X}_e,\, e \in \mathcal{E}, \tag{193}$$

where $\boldsymbol{x}' = (x_e')_{e \in \mathcal{E}}$ is given in (117), then we have

$$Z_{\mathrm{B,d,N}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) \overset{(a)}{\geq} g(\boldsymbol{x}') > 0, \qquad \boldsymbol{\lambda} \in \mathbb{R}^{|\mathcal{X}|}, \tag{194}$$

where step (a) follows from the definition of $Z_{\mathrm{B,d,N}}$ in (18), the definition of $F_{\mathrm{B,d,N}}^{\mathrm{alt}}$ in (23), and the definition of Z_f in (16), i.e.,

$$Z_{\mathrm{B,d,N}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) = \exp\big(-F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}, \boldsymbol{\lambda})\big) = \sum_{\{\boldsymbol{x}_{\partial f,f}\}_f} \prod_{f \in \mathcal{F}} \left(f(\boldsymbol{x}_{\partial f,f}) \prod_{e \in \partial f} \Big(\exp\big(\lambda_{e,f}(x_{e,f})\big) \cdot \sqrt{\gamma_e(x_{e,f})} \Big) \right).$$

The inequalities in (194) also imply

$$F_{\mathrm{B,d,N}}^{\mathrm{alt},*} \overset{(a)}{\leq} \hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}) \overset{(b)}{=} \sup_{\boldsymbol{\lambda} \in \mathbb{R}^{|\mathcal{X}|}} F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) \leq -\log g(\boldsymbol{x}') < \infty,$$

where step (a) follows from the definition of $F^{\mathrm{alt},*}_{\mathrm{B,d,N}}$ in Item 5 in Definition 12 and where step (b) follows from the definition of $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma})$ in (29). With this, we have

$$F^{\mathrm{alt},*}_{\mathrm{B,d,N}} \overset{(a)}{=} \lim_{m\to\infty} \hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}) \overset{(b)}{=} \lim_{m\to\infty} \sup_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \boldsymbol{\lambda}), \tag{195}$$

where step (a) follows from (30) and where step (b) follows from the definition of $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma})$ in (29). Then we have

$$\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}) = \sup_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \boldsymbol{\lambda}) \overset{(a)}{\geq} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \mathbf{0}) \overset{(b)}{\geq} -\sum_{f} \log\left(\sum_{\boldsymbol{x}_{\partial f,f}} f(\boldsymbol{x}_{\partial f,f})\right), \qquad m \in \mathbb{Z}_{>0}, \tag{196}$$

where step (a) follows from (31) and where step (b) follows from $\gamma_e^{(m)}(x_e) \in [0,1]$ for all $x_e \in \mathcal{X}_e$ and $e \in \mathcal{E}$ and $f(\boldsymbol{x}_{\partial f,f}) \in \mathbb{R}_{\geq 0}$ for all $\boldsymbol{x}_{\partial f,f} \in \mathcal{X}_f$ as defined in Definition 3, i.e.,

$$F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \mathbf{0}) = -\log\left(\prod_{f\in\mathcal{F}}\left(\sum_{\boldsymbol{x}_{\partial f,f}} f(\boldsymbol{x}_{\partial f,f}) \cdot \prod_{e\in\partial f} \sqrt{\gamma_e^{(m)}(x_{e,f})}\right)\right) \geq -\log\left(\prod_{f\in\mathcal{F}}\left(\sum_{\boldsymbol{x}_{\partial f,f}} f(\boldsymbol{x}_{\partial f,f})\right)\right)$$

Combining (195) with (196), we get

$$F^{\mathrm{alt},*}_{\mathrm{B,d,N}} = \lim_{m\to\infty} \sup_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \boldsymbol{\lambda}) \geq -\sum_{f} \log\left(\sum_{\boldsymbol{x}_{\partial f,f}} f(\boldsymbol{x}_{\partial f,f})\right) > -\infty.$$

Appendix O

Proof of Proposition 52

We prove the claims in this proposition item by item.

1) This follows from the definitions of $F^{\mathrm{alt}}_{\mathrm{B,d,N}}$ and Z_f in (23) and (16), respectively.
2) It can be proven by (118) in Proposition 51 and the property of $\boldsymbol{\gamma}^{(m)}$ in (30).
3) We prove each property separately.
   a) We prove it by contradiction. Suppose that there exists an $f' \in \mathcal{F}$ such that

   $$Z_{f'}\bigl(\boldsymbol{\gamma}^{(m)}_{\partial f'},\, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}_{\partial f'}\bigr) = 0.$$

   For fixed n and k, because $\boldsymbol{\gamma}_e^{(m)} \in \mathcal{B}_e^{\geq}$ as defined in Item 2 in Definition 2, the function Z_f is bounded and we have

   $$\prod_{f\in\mathcal{F}} Z_f\bigl(\boldsymbol{\gamma}^{(m)}_{\partial f},\, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}_{\partial f}\bigr) = 0.$$

   By the definition of $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}$ in (29), we have

   $$\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}) \geq -\log\left(\prod_{f} Z_f\bigl(\boldsymbol{\gamma}^{(m)}_{\partial f},\, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}_{\partial f}\bigr)\right) = \infty,$$

   which is a contradiction of the inequalities in (121).
   b) It can proven by the property in (122) and the definition of $b_f^{(m,n,k)}$ in (37).
   c) It can be proven by combining the inequalities in (121) with the inequalities in (119) as stated in Proposition 51.
   d) The proof of convergence follows from the definition of the sequence $\{\boldsymbol{\gamma}^{(m)}\}_{m\in\mathbb{Z}_{>0}}$ in Item 2 in Definition 17. The proof of $F^{\mathrm{alt},*}_{\mathrm{B,d,N}} \in \mathbb{R}$ follows from (118) in Proposition 51
   e) By inequalities in (121), we have

   $$\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}) \in \mathbb{R}.$$

   By the definition of the sequence $\{\boldsymbol{\lambda}^{(n)}\}_{n\in\mathbb{Z}_{>0}}$ in Item 3 in Definition 17, we can prove that

   $$\lim_{n\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \boldsymbol{\lambda}^{(n)}) = \hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}).$$

f) By the definition of $\{\alpha^{(k)}\}_k$ in Item 4 in Definition 17, we can prove the equality in (126). The proof of the property

$$\max_{\alpha\in\mathbb{R}\cup\{-\infty,+\infty\}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\alpha\cdot\boldsymbol{\lambda}^{(n)})\in\mathbb{R}$$

follows from

$$-\log g(\boldsymbol{x}')\overset{(a)}{\geq} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)})\overset{(b)}{\geq} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)})\overset{(c)}{\geq} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\mathbf{0})\in\mathbb{R},$$

where step (a) follows from the inequalities in (121), where step (b) follows from (29), and where step (c) follows from the definition of the sequence $\{\alpha^{(k)}\}_k$ in (35) in Definition 17.

Appendix P
Proof of Theorem 54

By the definition of $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)})$ in (29) and property (121) in Proposition 52, we know that

$$F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)})\leq\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)})\in\mathbb{R},\qquad n,k\in\mathbb{Z}_{>0}.$$

By the property of $\alpha^{(k)}$ in (126), we have

$$\lim_{k\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)})-F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\boldsymbol{\lambda}^{(n)})\geq 0,\qquad n\in\mathbb{Z}_{>0}. \tag{197}$$

As shown in (125), the sequence $\big\{F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\boldsymbol{\lambda}^{(n)})\big\}_{n\in\mathbb{Z}_{>0}}$ converges in $\mathbb{R}$, which means that this sequence is a bounded sequence. Then we know that the following sequences are bounded as well:

$$\Big\{\lim_{k\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)})\Big\}_{n\in\mathbb{Z}_{>0}},$$
$$\Big\{\lim_{k\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)})-F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\boldsymbol{\lambda}^{(n)})\Big\}_{n\in\mathbb{Z}_{>0}}.$$

1) It holds that

$$\begin{aligned}&\limsup_{n\to\infty}\Big(\lim_{k\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)})-F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\boldsymbol{\lambda}^{(n)})\Big)\\&\leq\limsup_{n\to\infty}\lim_{k\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)})-\liminf_{n\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\boldsymbol{\lambda}^{(n)})\\&\overset{(a)}{=}\limsup_{n\to\infty}\lim_{k\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)})-\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)})\\&\overset{(b)}{\leq} 0,\end{aligned}$$

where step (a) follows from the property of $\boldsymbol{\lambda}^{(n)}$ in (125) and where step (b) follows from the definition of $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)})$ in (29).

2) By (197), we have

$$\begin{aligned}0&\leq\liminf_{n\to\infty}\Big(\lim_{k\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)})-F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\boldsymbol{\lambda}^{(n)})\Big)\\&\overset{(a)}{=}\liminf_{n\to\infty}\lim_{k\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)})-\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}),\end{aligned}$$

where step (a) again follows from the property of $\boldsymbol{\lambda}^{(n)}$ in (125).

3) Thus we have

$$\begin{aligned}\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)})&\leq\liminf_{n\to\infty}\lim_{k\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)})\\&\leq\limsup_{n\to\infty}\lim_{k\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)})\\&\leq\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}).\end{aligned}$$

APPENDIX Q
PROOF OF LEMMA 56

For arbitrarily small $r \in \mathbb{R}_{>0}$ such that $|r| < 1$, we define $\boldsymbol{\delta_\lambda} \in \mathbb{R}^{|\mathcal{X}|}$ such that $\|\boldsymbol{\delta_\lambda}\|_2 = 1$ and $\boldsymbol{\delta_\lambda} = (\delta_{\boldsymbol{\lambda}_e}(x_e))_{x_e \in \mathcal{X}_e, e \in \mathcal{E}}$. For each $e = (f_i, f_j)$ such that $i < j$, we define

$$\delta_{\boldsymbol{\lambda}_{e,f_i}} = \delta_{\boldsymbol{\lambda}_e}, \qquad \delta_{\boldsymbol{\lambda}_{e,f_j}} = -\delta_{\boldsymbol{\lambda}_e},$$

which is consistent with the signs in the definition of $\boldsymbol{\lambda}_{e,f}$ in (15). We first prove that the following sequence is bounded:

$$\left\{ F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)} + r \cdot \boldsymbol{\delta_\lambda}) \right\}_{n,k \in \mathbb{Z}_{>0}}. \tag{198}$$

It holds that

$$\begin{aligned} 0 &\leq \prod_{f \in \mathcal{F}} \left(f(\boldsymbol{x}_f) \cdot \prod_{e \in \partial f} \left(\exp\left(\alpha^{(k)} \cdot \lambda^{(n)}_{e,f}(x_e) + r \cdot \delta_{\boldsymbol{\lambda}_{e,f}}(x_e)\right) \cdot \sqrt{\gamma^{(m)}_e(x_e)} \right) \right) \\ &\overset{(a)}{\leq} \prod_{f \in \mathcal{F}} \left(f(\boldsymbol{x}_f) \cdot \prod_{e \in \partial f} \left(\exp\left(\alpha^{(k)} \cdot \lambda^{(n)}_{e,f}(x_e)\right) \cdot \exp(1) \cdot \sqrt{\gamma^{(m)}_e(x_e)} \right) \right) \\ &\overset{(b)}{\leq} \exp\left(-F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \mathbf{0}) + 2|\mathcal{E}| \right), \qquad \boldsymbol{x}_f \in \mathcal{X}_f,\ f \in \mathcal{F},\ \forall n, k \in \mathbb{Z}_{>0}, \end{aligned} \tag{199}$$

where step (a) follows from

$$|r \cdot \delta_{\boldsymbol{\lambda}_{e,f}}(x_e)| = |r \cdot \delta_{\boldsymbol{\lambda}_e}(x_e)| \leq 1, \qquad x_e \in \mathcal{X}_e,\ e \in \mathcal{E},$$

where step (b) follows from inequality (128) in Proposition 53 and $\sum_f \sum_{e \in \partial f} 1 = 2|\mathcal{E}|$. Then by (120) we get

$$\begin{aligned} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)} + r \cdot \boldsymbol{\delta_\lambda}) &= -\log\left(\sum_{\{\boldsymbol{x}_{\partial f,f}\}_f} \prod_{f \in \mathcal{F}} \left(f(\boldsymbol{x}_{\partial f,f}) \cdot \prod_{e \in \partial f} \left(\exp\left(\alpha^{(k)} \cdot \lambda^{(n)}_{e,f}(x_{e,f}) + r \cdot \delta_{\boldsymbol{\lambda}_{e,f}}(x_{e,f})\right) \cdot \sqrt{\gamma^{(m)}_e(x_{e,f})} \right) \right) \right) \\ &\overset{(a)}{\geq} -\log\left(\sum_f |\mathcal{X}_f| \right) + F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \mathbf{0}) - 2|\mathcal{E}|, \qquad n, k \in \mathbb{Z}_{>0}, \end{aligned}$$

where step (a) follows from (199) and $\sum_{\{\boldsymbol{x}_{\partial f,f}\}_f} 1 = \sum_f |\mathcal{X}_f|$. Note that by inequalities (123) in Proposition 52, we have

$$F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \mathbf{0}) \in \mathbb{R}, \qquad m \in \mathbb{Z}_{>0}.$$

By the definition of $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}$ in (29) and the property $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}) \in \mathbb{R}$ as proven in (125), we know that

$$F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)} + r \cdot \boldsymbol{\delta_\lambda}) \leq \hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}) \in \mathbb{R}, \qquad n, k \in \mathbb{Z}_{>0}.$$

Now we know that the sequence in (198) is bounded. Combining with Theorem 54, we know that the following limits exist

$$\liminf_{n\to\infty} \liminf_{k\to\infty} \left(F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}) - F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)} + r \cdot \boldsymbol{\delta_\lambda}) \right), \quad \limsup_{n\to\infty} \limsup_{k\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)} + r \cdot \boldsymbol{\delta_\lambda}).$$

Then we want to use the Taylor series expansion of $F^{\mathrm{alt}}_{\mathrm{B,d,N}}$ to prove this lemma. It holds that

$$\begin{aligned} &\liminf_{n\to\infty} \liminf_{k\to\infty} \left(F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}) - F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)} + r \cdot \boldsymbol{\delta_\lambda}) \right) \\ &\geq \liminf_{n\to\infty} \liminf_{k\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}) - \limsup_{n\to\infty} \limsup_{k\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)} + r \cdot \boldsymbol{\delta_\lambda}) \\ &\overset{(a)}{=} \hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}) - \limsup_{n\to\infty} \limsup_{k\to\infty} F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)} + r \cdot \boldsymbol{\delta_\lambda}) \\ &\overset{(b)}{\geq} 0, \end{aligned} \tag{200}$$

where step (a) follows Theorem 54 and where step (b) follows from the definition of $\hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}$ in (29). The Taylor series expansion of $F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)} + r \cdot \boldsymbol{\delta}_{\boldsymbol{\lambda}})$ w.r.t. r at $r = 0$ is given by

$$F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)} + r \cdot \boldsymbol{\delta}_{\boldsymbol{\lambda}}) = F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}) + r \cdot \left(\nabla_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}\big|_{\boldsymbol{\lambda}=\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)},\, \boldsymbol{\gamma}=\boldsymbol{\gamma}^{(m)}} \right)^{\mathsf{T}} \cdot \boldsymbol{\delta}_{\boldsymbol{\lambda}}$$
$$+ \frac{r^2}{2} \cdot \boldsymbol{\delta}_{\boldsymbol{\lambda}}^{\mathsf{T}} \cdot \left(\nabla^2_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}\big|_{\boldsymbol{\lambda}=\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)}+r'\cdot\boldsymbol{\delta}_{\boldsymbol{\lambda}},\, \boldsymbol{\gamma}=\boldsymbol{\gamma}^{(m)}} \right) \cdot \boldsymbol{\delta}_{\boldsymbol{\lambda}}, \tag{201}$$

where $r' \in [0, r]$ is suitably chosen, the vector $\nabla_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}$ is the gradient of $F^{\mathrm{alt}}_{\mathrm{B,d,N}}$ w.r.t. $\boldsymbol{\lambda}$ with entries given by

$$\frac{\partial}{\partial \lambda_e(x_e)} F^{\mathrm{alt}}_{\mathrm{B,d,N}} = -b_{f_i,e}(x_e) + b_{f_j,e}(x_e), \tag{202}$$

the function $b_{f,e}$ is given in (25), and the matrix $\nabla^2_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}$ is the Hessian matrix of $F^{\mathrm{alt}}_{\mathrm{B,d,N}}$ w.r.t. $\boldsymbol{\lambda}$. By the entries of the gradient $\nabla_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}$ in (202), the entries in the Hessian matrix $\nabla^2_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}$ are given by

$$\frac{\partial}{\partial \lambda_{e_2}(x_{e_2})} \Big(\frac{\partial}{\partial \lambda_{e_1}(x_{e_1})} F^{\mathrm{alt}}_{\mathrm{B,d,N}} \Big) = -\frac{\partial}{\partial \lambda_{e_2}(x_{e_2})} b_{f_i,e_1}(x_{e_1}) + \frac{\partial}{\partial \lambda_{e_2}(x_{e_2})} b_{f_j,e_1}(x_{e_1}),$$
$$e_1 = (f_i, f_j),\, i < j,\, x_{e_1} \in \mathcal{X}_{e_1}, \qquad e_2 = (f_k, f_\ell),\, k < \ell,\, x_{e_2} \in \mathcal{X}_{e_2},$$

$$\frac{\partial}{\partial \lambda_{e_2}(x_{e_2})} b_{f_i,e_1}(x_{e_1}) = [i \in \{k, \ell\}] \cdot \begin{cases} b_{f_i,\{e_1,e_2\}}(x_{e_1}, x_{e_2}) - b_{f_i,e_1}(x_{e_1}) \cdot b_{f_k,e_2}(x_{e_2}) & e_1 \neq e_2,\, k = i \\ b_{f_i,e_1}(x_{e_1}) \cdot b_{f_\ell,e_2}(x_{e_2}) - b_{f_i,\{e_1,e_2\}}(x_{e_1}, x_{e_2}) & e_1 \neq e_2,\, \ell = i,\, , \\ [x_{e_1} = x_{e_2}] \cdot b_{f_i,e_1}(x_{e_1}) - b_{f_i,e_1}(x_{e_1}) \cdot b_{f_k,e_2}(x_{e_2}) & e_1 = e_2, \end{cases} \tag{203}$$

$$\frac{\partial}{\partial \lambda_{e_2}(x_{e_2})} b_{f_j,e_1}(x_{e_1}) = [j \in \{k, \ell\}] \cdot \begin{cases} b_{f_j,\{e_1,e_2\}}(x_{e_1}, x_{e_2}) - b_{f_j,e_1}(x_{e_1}) \cdot b_{f_k,e_2}(x_{e_2}) & e_1 \neq e_2,\, k = j, \\ b_{f_j,e_1}(x_{e_1}) \cdot b_{f_\ell,e_2}(x_{e_2}) - b_{f_j,\{e_1,e_2\}}(x_{e_1}, x_{e_2}) & e_1 \neq e_2,\, \ell = j, \\ -[x_{e_1} = x_{e_2}] \cdot b_{f_j,e_1}(x_{e_1}) + b_{f_i,e_1}(x_{e_1}) \cdot b_{f_k,e_2}(x_{e_2}) & e_1 = e_2, \end{cases} \tag{204}$$

where the marginals of b_f are defined in (39). The above expressions and the fact that $0 \leq b_f(\boldsymbol{x}_f) \leq 1$ for all $\boldsymbol{x}_f \in \mathcal{X}_f,\ f \in \mathcal{F}$, imply that the following sequences are bounded for $\|\boldsymbol{\delta}_{\boldsymbol{\lambda}}\|_2 = 1$:

$$\left\{ \boldsymbol{\delta}_{\boldsymbol{\lambda}}^{\mathsf{T}} \cdot \left(\nabla^2_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}\big|_{\boldsymbol{\lambda}=\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)}+r'\cdot\boldsymbol{\delta}_{\boldsymbol{\lambda}},\, \boldsymbol{\gamma}=\boldsymbol{\gamma}^{(m)}} \right) \cdot \boldsymbol{\delta}_{\boldsymbol{\lambda}} \right\}_{n,k}, \qquad \left\{ -\left(\nabla_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}\big|_{\boldsymbol{\lambda}=\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)},\, \boldsymbol{\gamma}=\boldsymbol{\gamma}^{(m)}} \right)^{\mathsf{T}} \cdot \boldsymbol{\delta}_{\boldsymbol{\lambda}} \right\}_{n,k}.$$

Thus we have

$$\begin{aligned}
0 &\overset{(a)}{\leq} \liminf_{r\downarrow 0} \liminf_{n\to\infty} \liminf_{k\to\infty} \frac{1}{r} \Big(F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)}) - F^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}, \alpha^{(k)} \cdot \boldsymbol{\lambda}^{(n)} + r \cdot \boldsymbol{\delta}_{\boldsymbol{\lambda}}) \Big) \\
&\overset{(b)}{=} \liminf_{r\downarrow 0} \liminf_{n\to\infty} \liminf_{k\to\infty} \left(-\nabla_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}\big|_{\boldsymbol{\lambda}=\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)},\, \boldsymbol{\gamma}=\boldsymbol{\gamma}^{(m)}} \cdot \boldsymbol{\delta}_{\boldsymbol{\lambda}} - \frac{r}{2} \cdot \boldsymbol{\delta}_{\boldsymbol{\lambda}}^{\mathsf{T}} \cdot \left(\nabla^2_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}\big|_{\boldsymbol{\lambda}=\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)}+r'\cdot\boldsymbol{\delta}_{\boldsymbol{\lambda}},\, \boldsymbol{\gamma}=\boldsymbol{\gamma}^{(m)}} \right) \cdot \boldsymbol{\delta}_{\boldsymbol{\lambda}} \right) \\
&\leq \liminf_{n\to\infty} \liminf_{k\to\infty} -\nabla_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}\big|_{\boldsymbol{\lambda}=\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)},\, \boldsymbol{\gamma}=\boldsymbol{\gamma}^{(m)}} \cdot \boldsymbol{\delta}_{\boldsymbol{\lambda}} - \limsup_{r\downarrow 0} \frac{r}{2} \cdot \underbrace{\limsup_{n\to\infty} \limsup_{k\to\infty} \boldsymbol{\delta}_{\boldsymbol{\lambda}}^{\mathsf{T}} \cdot \left(\nabla^2_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}\big|_{\boldsymbol{\lambda}=\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)}+r'\cdot\boldsymbol{\delta}_{\boldsymbol{\lambda}},\, \boldsymbol{\gamma}=\boldsymbol{\gamma}^{(m)}} \right) \cdot \boldsymbol{\delta}_{\boldsymbol{\lambda}}}_{\text{is bounded } \forall r > 0.} \\
&\overset{(c)}{=} \liminf_{n\to\infty} \liminf_{k\to\infty} -\nabla_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}\big|_{\boldsymbol{\lambda}=\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)},\, \boldsymbol{\gamma}=\boldsymbol{\gamma}^{(m)}} \cdot \boldsymbol{\delta}_{\boldsymbol{\lambda}},
\end{aligned} \tag{205}$$

where step (a) follows from (200) and $r \in \mathbb{R}_{>0}$, where step (b) follows from (201), where step (c) follows from the fact that a product of a real-valued number and a sequence that converges to zero is again a sequence converging to zero. Note that the inequalities in (205) hold for any $\boldsymbol{\delta}_{\boldsymbol{\lambda}} \in \mathbb{R}^{|\mathcal{X}|}$ such that $\|\boldsymbol{\delta}_{\boldsymbol{\lambda}}\| = 1$. We have

$$\lim_{n,k\to\infty} \nabla_{\boldsymbol{\lambda}} F^{\mathrm{alt}}_{\mathrm{B,d,N}}\big|_{\boldsymbol{\lambda}=\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)},\, \boldsymbol{\gamma}=\boldsymbol{\gamma}^{(m)}} = \mathbf{0},$$

which by (202), implies

$$\lim_{n,k\to\infty}\Big(-b_{f_i,e}^{(m,n,k)}(x_e)+b_{f_j,e}^{(m,n,k)}(x_e)\Big)=0,\qquad \boldsymbol{\gamma}=\boldsymbol{\gamma}^{(m)},\ x_e\in\mathcal{X}_e,\ e=(f_i,f_j)\in\mathcal{E}.$$

APPENDIX R
PROOF OF THEOREM 57

Before proving this theorem, we prove two lemmas.

Lemma 63. *It holds that*

$$\lim_{k\to\infty}\alpha^{(k)}\sum_{f}\sum_{\boldsymbol{x}_f} b_f^{(m,n,k)}(\boldsymbol{x}_f)\cdot\left(\sum_{e\in\partial f}\lambda_{e,f}^{(n)}(x_{e,f})\right)=0,\qquad m,n\in\mathbb{Z}_{>0},$$

where $\boldsymbol{b}_f^{(m,n,k)}$ *is defined in* (37).

Proof. If we can prove that

$$\lim_{k\to\infty}\alpha^{(k)}\cdot\frac{\partial}{\partial\alpha}F_{\mathrm{B,d,N}}^{\mathrm{alt}}\Big|_{\alpha=\alpha^{(k)}}=\lim_{k\to\infty}\alpha^{(k)}\cdot\left(\nabla_{\boldsymbol{\lambda}}F_{\mathrm{B,d,N}}^{\mathrm{alt}}\big|_{\boldsymbol{\lambda}=\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)},\,\boldsymbol{\gamma}=\boldsymbol{\gamma}^{(m)}}\right)^{\mathsf{T}}\cdot\boldsymbol{\lambda}^{(n)}=0,\qquad m,n\in\mathbb{Z}_{>0},\tag{206}$$

where $\nabla_{\boldsymbol{\lambda}}F_{\mathrm{B,d,N}}^{\mathrm{alt}}$ is the gradient of $F_{\mathrm{B,d,N}}^{\mathrm{alt}}$ w.r.t. $\boldsymbol{\lambda}$, and the entries in the gradient are given in (202). Then we have

$$\begin{aligned}
&\lim_{k\to\infty}\alpha^{(k)}\cdot\left(\nabla_{\boldsymbol{\lambda}}F_{\mathrm{B,d,N}}^{\mathrm{alt}}\big|_{\boldsymbol{\lambda}=\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)},\,\boldsymbol{\gamma}=\boldsymbol{\gamma}^{(m)}}\right)^{\mathsf{T}}\cdot\boldsymbol{\lambda}^{(n)}\\
&=\lim_{k\to\infty}\alpha^{(k)}\sum_{e=(f_i,f_j)}\sum_{x_e}\lambda_e^{(n)}(x_e)\cdot\big(-b_{f_i,e}^{(m,n,k)}(x_e)+b_{f_j,e}^{(m,n,k)}(x_e)\big)\\
&\overset{(a)}{=}\lim_{k\to\infty}-\alpha^{(k)}\sum_{f}\sum_{\boldsymbol{x}_f}b_f^{(m,n,k)}(\boldsymbol{x}_f)\sum_{e\in\partial f}\lambda_{e,f}^{(n)}(x_{e,f})\\
&=0,\qquad n,m\in\mathbb{Z}_{>0},
\end{aligned}$$

where step (a) follows from the definition of $\{\boldsymbol{\lambda}_{e,f}\}_{e\in\partial f,f\in\mathcal{F}}$ in (15). Then we prove the lemma.

In the following, we prove (206). There are various cases that need to be discussed.

1) Suppose that $\alpha^*=\infty$ for some m and n, if

$$\exists\{\boldsymbol{x}_{\partial f,f}\}_f \text{ such that } \prod_{f\in\mathcal{F}}\left(f(\boldsymbol{x}_{\partial f,f})\cdot\prod_{e\in\partial f}\sqrt{\gamma_e^{(m)}(x_{e,f})}\right)>0,\qquad \sum_{e=(f_i,f_j)}\Big(\lambda_e^{(n)}(x_{e,f_i})-\lambda_e^{(n)}(x_{e,f_j})\Big)>0,$$

then we have

$$\begin{aligned}
&\lim_{k\to\infty}F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)},\alpha^{(k)}\cdot\boldsymbol{\lambda}^{(n)})\\
&=\lim_{k\to\infty}-\log\left(\prod_f Z_f\big(\boldsymbol{\gamma}_{\partial f}^{(m)},\,\alpha^{(k)}\cdot\boldsymbol{\lambda}_{\partial f}^{(n)}\big)\right)\\
&\overset{(a)}{=}\lim_{k\to\infty}-\log\sum_{\{\boldsymbol{x}_{\partial f,f}\}_f}\prod_{f\in\mathcal{F}}\left(f(\boldsymbol{x}_{\partial f,f})\cdot\prod_{e\in\partial f}\left(\exp\big(\alpha^{(k)}\cdot\lambda_{e,f}^{(n)}(x_{e,f})\big)\cdot\sqrt{\gamma_e^{(m)}(x_{e,f})}\right)\right)\\
&=\lim_{k\to\infty}-\log\left(\sum_{\{\boldsymbol{x}_{\partial f,f}\}_f}\left(\prod_{f\in\mathcal{F}}f(\boldsymbol{x}_{\partial f,f})\cdot\prod_{e\in\partial f}\sqrt{\gamma_e^{(m)}(x_{e,f})}\right)\cdot\exp\left(\alpha^{(k)}\sum_{e=(f_i,f_j)}\big(\lambda_e^{(n)}(x_{e,f_i})-\lambda_e^{(n)}(x_{e,f_j})\big)\right)\right)\\
&=-\infty\\
&\overset{(b)}{<}-\log\left(\prod_f Z_f\big(\boldsymbol{\gamma}_{\partial f}^{(m)},\mathbf{0}\big)\right)
\end{aligned}$$

$$\stackrel{(c)}{=} F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}, \mathbf{0}),$$

where step (a) follows from (120), where step (b) follows from the inequalities in (123), i.e.,

$$F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}, \mathbf{0}) = -\log\left(\prod_{f} Z_f\big(\boldsymbol{\gamma}_{\partial f}^{(m)}, \mathbf{0}\big)\right) \in \mathbb{R},$$

and where step (c) follows from the definition of $F_{\mathrm{B,d,N}}^{\mathrm{alt}}$ in (23). The above inequalities contradict the properties of $\{\alpha^{(k)}\}$ in (35) and (126). Thus we only need to consider the following terms in $\prod_f Z_f$:

$$\prod_{f\in\mathcal{F}}\left(f(\boldsymbol{x}_{\partial f,f}) \cdot \prod_{e\in\partial f}\sqrt{\gamma_e^{(m)}(x_{e,f})}\right) > 0, \qquad \sum_{e=(f_i,f_j)}\Big(\lambda_e^{(n)}(x_{e,f_i}) - \lambda_e^{(n)}(x_{e,f_j})\Big) \leq 0, \quad \boldsymbol{x}_{\partial f,f} \in \mathcal{X}_f,\ f \in \mathcal{F}. \tag{207}$$

By (126), we know that there exists $\{\boldsymbol{x}_{\partial f,f}^{*}\}_f$ such that

$$\prod_{f\in\mathcal{F}}\left(f(\boldsymbol{x}_{\partial f,f}^{*}) \cdot \prod_{e\in\partial f}\sqrt{\gamma_e^{(m)}(x_{e,f}^{*})}\right) > 0, \qquad \sum_{e=(f_i,f_j)}\Big(\lambda_{e,f_i}^{(n)}(x_{e,f_i}^{*}) - \lambda_{e,f_j}^{(n)}(x_{e,f_j}^{*})\Big) = 0. \tag{208}$$

Otherwise, by (120), we have $\sup_{\alpha\in\mathbb{R}} F_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}, \alpha\cdot\boldsymbol{\lambda}^{(n)}) = \infty$, which is a contradiction of (126). Then it holds that

$$\prod_{f} Z_f\big(\boldsymbol{\gamma}_{\partial f}^{(m)},\, \mathbf{0}\big) \stackrel{(a)}{\geq} \prod_{f} Z_f\big(\boldsymbol{\gamma}_{\partial f}^{(m)},\, \alpha^{(k)}\cdot\boldsymbol{\lambda}_{\partial f}^{(n)}\big) \stackrel{(b)}{\geq} \prod_{f\in\mathcal{F}}\left(f(\boldsymbol{x}_{\partial f,f}^{*}) \cdot \prod_{e\in\partial f}\sqrt{\gamma_e^{(m)}(x_{e,f}^{*})}\right) > 0, \qquad k \in \mathbb{Z}_{>0},$$

where step (a) follows from the definitions of $F_{\mathrm{B,d,N}}^{\mathrm{alt}}$ and $\alpha^{(k)}$ in (23) and (35), respectively and where step (b) follows from the expression of $-\log\prod_f Z_f$ in (120) and the expressions in (208). Then for each n, there exists $M_1 \in \mathbb{R}_{>0}$ such that the following inequalities hold for fixed $\boldsymbol{\gamma}^{(m)}$ and $\boldsymbol{\lambda}^{(n)}$:

$$0 \leq \sum_{\{\boldsymbol{x}_{\partial f,f}\}_f}\prod_{f}\frac{f(\boldsymbol{x}_{\partial f,f})\cdot\prod_{e\in\partial f}\sqrt{\gamma_e^{(m)}(x_{e,f})}}{Z_f\big(\boldsymbol{\gamma}_{\partial f}^{(m)},\, \alpha^{(k)}\cdot\boldsymbol{\lambda}_{\partial f}^{(n)}\big)}\cdot\left(\sum_{e=(f_i,f_j)}\Big(\lambda_e^{(n)}(x_{e,f_i}) - \lambda_e^{(n)}(x_{e,f_j})\Big)\right) \leq M_1, \qquad k \in \mathbb{Z}_{>0}.$$

Then by $\lim_{k\to\infty}\alpha^{(k)}\cdot\exp\!\big(c\cdot\alpha^{(k)}\big) = 0$ for $c < 0$ and the expression $F_{\mathrm{B,d,N}}^{\mathrm{alt}}$ in (120), we have

$$\begin{aligned}
\lim_{k\to\infty}\alpha^{(k)}\cdot\frac{\partial}{\partial\alpha}F_{\mathrm{B,d,N}}^{\mathrm{alt}}\bigg|_{\alpha=\alpha^{(k)}} &= -\lim_{k\to\infty}\sum_{\{\boldsymbol{x}_{\partial f,f}\}_f}\underbrace{\prod_{f}\frac{f(\boldsymbol{x}_{\partial f,f})\cdot\prod_{e\in\partial f}\Big(\sqrt{\gamma_e^{(m)}(x_{e,f})}\Big)}{Z_f\big(\boldsymbol{\gamma}_{\partial f}^{(m)},\, \alpha^{(k)}\cdot\boldsymbol{\lambda}_{\partial f}^{(n)}\big)}\cdot\left(\sum_{e=(f_i,f_j)}\Big(\lambda_e^{(n)}(x_{e,f_i}) - \lambda_e^{(n)}(x_{e,f_j})\Big)\right)}_{\text{is bounded for fixed } k} \\
&\qquad\cdot\alpha^{(k)}\cdot\exp\left(\alpha^{(k)}\sum_{e=(f_i,f_j)}\big(\lambda_e^{(n)}(x_{e,f_i}) - \lambda_e^{(n)}(x_{e,f_j})\big)\right) \\
&= 0.
\end{aligned}$$

2) Suppose that $\alpha^{*} = -\infty$ for some n, similar to the proof in the previous case, we have

$$\lim_{k\to\infty}\alpha^{(k)}\cdot\frac{\partial}{\partial\alpha}F_{\mathrm{B,d,N}}^{\mathrm{alt}}\bigg|_{\alpha=\alpha^{(k)}} = 0.$$

3) Suppose that $\alpha^{*} \in \mathbb{R}$, we have $\alpha^{(k)} = \alpha^{*}$ for all k. The location of the optimal value for the optimization problem in (33) satisfies

$$\alpha^{*}\cdot\frac{\partial}{\partial\alpha}F_{\mathrm{B,d,N}}^{\mathrm{alt}}\bigg|_{\alpha=\alpha^{*}} = 0.$$

■

Lemma 64. *For fixed $\boldsymbol{\gamma}^{(m)}$, the following sequence is bounded*

$$\Big\{F_{\mathrm{B,p,N}}^{(1)}\big(\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}\big)\Big\}_{n,k\in\mathbb{Z}_{>0}},$$

where $\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)} = (\boldsymbol{b}_{f}^{(m,n,k)})_f$ *as defined in* (38).

Proof. As defined in (37), we have $0 \leq b_f^{(m,n,k)}(\boldsymbol{x}_f) \leq 1$ for all $\boldsymbol{x}_f \in \mathcal{X}_f$ and $f \in \mathcal{F}$. It holds that

$$\left| \sum_{e=(f_i,f_j)\in\mathcal{E}} \sum_{x_e:\, \gamma_e^{(m)}(x_e)>0} \frac{b_{f_i,e}^{(m,n,k)}(x_e) + b_{f_j,e}^{(m,n,k)}(x_e)}{2} \cdot \log \gamma_e^{(m)}(x_e) \right| \leq \sum_{e} \sum_{x_e:\, \gamma_e^{(m)}(x_e)>0} \left| \log \gamma_e^{(m)}(x_e) \right| < \infty, \qquad n,k \in \mathbb{Z}_{>0}.$$

Because the entropy function is finite for probability distributions with finite support, it holds that

$$\left| \sum_{e=(f_i,f_j)\in\mathcal{E}} \sum_{x_e:\, \gamma_e^{(m)}(x_e)>0} \frac{b_{f_i,e}^{(m,n,k)}(x_e) + b_{f_j,e}^{(m,n,k)}(x_e)}{2} \cdot \log\left(\frac{b_{f_i,e}^{(m,n,k)}(x_e) + b_{f_j,e}^{(m,n,k)}(x_e)}{2} \right) \right| \leq \sum_{e} \log |\mathcal{X}_e|, \qquad n,k \in \mathbb{Z}_{>0}.$$

Combining with Theorem 54 and Lemma 63, we know that the first two terms in (129) are bounded and thus the function $F_{\mathrm{B,p,N}}^{(1)}(\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)})$ given in (129) is bounded. ■

In the remaining part of this appendix, we turn to prove Theorem 57. The proof of (130) is divided into three parts.

1) By Lemma 64, we know that the following limit exists

$$\limsup_{n\to\infty} \limsup_{k\to\infty} F_{\mathrm{B,p,N}}^{(1)}(\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}).$$

By the expression of $F_{\mathrm{B,p,N}}^{(1)}$ in (129) and the proof of Lemma 64, we know that each term in $F_{\mathrm{B,p,N}}^{(1)}(\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)})$ is bounded for all $n,k \in \mathbb{Z}_{>0}$, where $\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)} = (\boldsymbol{b}_{f}^{(m,n,k)})_f$ as defined in (38). Then we have

$$\begin{aligned}
&\limsup_{n\to\infty} \limsup_{k\to\infty} F_{\mathrm{B,p,N}}^{(1)}(\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}) \\
&\leq \limsup_{n\to\infty} \limsup_{k\to\infty} -\log\left(\prod_f Z_f\big(\boldsymbol{\gamma}_{\partial f}^{(m)},\, \alpha^{(k)} \cdot \boldsymbol{\lambda}_{\partial f}^{(n)}\big) \right) \\
&\quad + \limsup_{n\to\infty} \limsup_{k\to\infty} \alpha^{(k)} \cdot \sum_f \sum_{\boldsymbol{x}_f} b_f^{(m,n,k)}(\boldsymbol{x}_f) \cdot \sum_{e\in\partial f} \lambda_{e,f}^{(n)}(x_{e,f}) \\
&\quad + \limsup_{n\to\infty} \limsup_{k\to\infty} \sum_{e=(f_i,f_j)\in\mathcal{E}} \sum_{x_e:\, \gamma_e^{(m)}(x_e)>0} \frac{b_{f_i,e}^{(m,n,k)}(x_e) + b_{f_j,e}^{(m,n,k)}(x_e)}{2} \log \frac{2\gamma_e^{(m)}(x_e)}{b_{f_i}^{(n)}(x_e) + b_{f_j}^{(n)}(x_e)} \\
&\overset{(a)}{=} \hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}) + \limsup_{n\to\infty} \limsup_{k\to\infty} \sum_{e=(f_i,f_j)\in\mathcal{E}} \sum_{x_e:\, \gamma_e^{(m)}(x_e)>0} \frac{b_{f_i,e}^{(m,n,k)}(x_e) + b_{f_j,e}^{(m,n,k)}(x_e)}{2} \log \frac{2\gamma_e^{(m)}(x_e)}{b_{f_i,e}^{(m,n,k)}(x_e) + b_{f_j,e}^{(m,n,k)}(x_e)} \qquad (209) \\
&\overset{(b)}{\leq} \hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}), \qquad (210)
\end{aligned}$$

where step (a) follows from Theorem 54 and Lemma 63 and where step (b) follows from the following conditions:

- $\boldsymbol{\gamma}_e^{(m)} \in \mathcal{B}_e^{\geq}$ for $e \in \mathcal{E}$, where $\mathcal{B}_e^{\geq}$ is defined in (5);
- it holds that

$$\boldsymbol{b}_f^{(m,n,k)} \in \mathcal{B}_f, \qquad f \in \mathcal{F}; \qquad (211)$$

- $(1/2) \cdot \sum_{x_e} \big(b_{f_i,e}^{(m,n,k)}(x_e) + b_{f_j,e}^{(m,n,k)}(x_e) \big) = 1$ for all $e = (f_i, f_j) \in \mathcal{E}$;
- if $\gamma_e^{(m)}(x_e) = 0$, then both $b_{f_i,e}^{(m,n,k)}(x_e)$ and $b_{f_j,e}^{(m,n,k)}(x_e)$ are zero;
- the Kullback–Leibler (K-L) divergence is nonnegative.

By (121) in Proposition 52, we know that $\hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)}) \in \mathbb{R}$, i.e., $\hat{F}_{\mathrm{B,d,N}}^{\mathrm{alt}}(\boldsymbol{\gamma}^{(m)})$ is bounded for fixed m.

2) By (211), we know that $\{b_f^{(m,n,k)}(\boldsymbol{x}_f)\}_{n,k\in\mathbb{Z}_{>0}}$ is a bounded sequence. By the Bolzano–Weierstrass theorem [23, Theorem 3.4.8], we can find a subsequence $\{n_1, k_1\}$ of $\{n, k\}$ such that the subsequence $\{b_f^{(m,n_1,k_1)}(\boldsymbol{x}_f)\}_{n_1,\,k_1\in\mathbb{Z}_{>0}}$ converges for all $\boldsymbol{x}_f$, $f \in \mathcal{F}$. By Lemma 56, we know that

$$\begin{aligned}\lim_{n_1\to\infty}\lim_{k_1\to\infty} b_{f_j,e}^{(m,n_1,k_1)}(x_e) &= \lim_{n_1\to\infty}\lim_{k_1\to\infty} b_{f_i,e}^{(m,n_1,k_1)}(x_e) - \lim_{n_1\to\infty}\lim_{k_1\to\infty}(b_{f_i,e}^{(m,n_1,k_1)}(x_e) - b_{f_j,e}^{(m,n_1,k_1)}(x_e)) \\ &= \lim_{n_1\to\infty}\lim_{k_1\to\infty} b_{f_i,e}^{(m,n_1,k_1)}(x_e), \qquad x_e\in\mathcal{X}_e,\, e=(f_i,f_j)\in\mathcal{E}.\end{aligned} \tag{212}$$

Combining with (211), we know that $\{\boldsymbol{b}_f^{(m,n_1,k_1)}\}_{n_1,k_1\in\mathbb{Z}_{>0}}$ converges to a point in $\mathcal{B}_{\mathcal{F}}(\mathsf{N})$. It holds that

$$\begin{aligned}F^*_{\mathrm{B,d},\mathsf{N}} &\overset{(a)}{=} F^*_{\mathrm{B,p},\mathsf{N}} \\ &\overset{(b)}{\leq} F^{(1)}_{\mathrm{B,p},\mathsf{N}}\Big(\lim_{n_1\to\infty}\lim_{k_1\to\infty}\boldsymbol{b}_{\mathcal{F}}^{(m,n_1,k_1)}\Big) \\ &\overset{(c)}{=} \lim_{n_1\to\infty}\lim_{k_1\to\infty} F^{(1)}_{\mathrm{B,p},\mathsf{N}}(\boldsymbol{b}_{\mathcal{F}}^{(m,n_1,k_1)}) \\ &\leq \limsup_{n\to\infty}\limsup_{k\to\infty} F^{(1)}_{\mathrm{B,p},\mathsf{N}}(\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}) \\ &\overset{(d)}{\leq} \hat{F}^{\mathrm{alt}}_{\mathrm{B,d},\mathsf{N}}(\boldsymbol{\gamma}^{(m)}),\end{aligned} \tag{213}$$

where step (a) follows from Proposition 16, where step (b) follows from the definition of $F^*_{\mathrm{B,p},\mathsf{N}}$ in (13) and the fact that $\{\boldsymbol{b}_{\mathcal{F}}^{(m,n_1,k_1)}\}_{n_1,k_1\in\mathbb{Z}_{>0}}$ converges to a point in $\mathcal{B}_{\mathcal{F}}(\mathsf{N})$, where step (c) follows from the continuity of $F^{(1)}_{\mathrm{B,p},\mathsf{N}}(\boldsymbol{b}_{\mathcal{F}})$ w.r.t. $\boldsymbol{b}_{\mathcal{F}}$ for each $\boldsymbol{b}_f\in\mathcal{B}_f$ and $f\in\mathcal{F}$, and where step (d) follows from (210).

3) By the property of $\boldsymbol{\gamma}^{(m)}$ in (124), we have

$$\lim_{m\to\infty}\hat{F}^{\mathrm{alt}}_{\mathrm{B,d},\mathsf{N}}(\boldsymbol{\gamma}^{(m)}) = F^{\mathrm{alt},*}_{\mathrm{B,d},\mathsf{N}} \overset{(a)}{\leq} F^*_{\mathrm{B,d},\mathsf{N}},$$

where step (a) follows from Proposition 15. Inequalities in (213) imply

$$\begin{aligned}F^*_{\mathrm{B,d},\mathsf{N}} &\leq \lim_{m,n_1,k_1\to\infty} F^{(1)}_{\mathrm{B,p},\mathsf{N}}(\boldsymbol{b}_{\mathcal{F}}^{(m,n_1,k_1)}) \\ &\leq \lim_{m\to\infty}\limsup_{n\to\infty}\limsup_{k\to\infty} F^{(1)}_{\mathrm{B,p},\mathsf{N}}(\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}) \\ &\leq \lim_{m\to\infty}\hat{F}^{\mathrm{alt}}_{\mathrm{B,d},\mathsf{N}}(\boldsymbol{\gamma}^{(m)}) \\ &= F^{\mathrm{alt},*}_{\mathrm{B,d},\mathsf{N}} \\ &\leq F^*_{\mathrm{B,d},\mathsf{N}}.\end{aligned} \tag{214}$$

Thus we have

$$\lim_{m,n_1,k_1\to\infty} F^{(1)}_{\mathrm{B,p},\mathsf{N}}(\boldsymbol{b}_{\mathcal{F}}^{(m,n_1,k_1)}) = \lim_{m\to\infty}\limsup_{n\to\infty}\limsup_{k\to\infty} F^{(1)}_{\mathrm{B,p},\mathsf{N}}(\boldsymbol{b}_{\mathcal{F}}^{(m,n,k)}) = F^{\mathrm{alt},*}_{\mathrm{B,d},\mathsf{N}} = F^*_{\mathrm{B,d},\mathsf{N}} \overset{(a)}{=} F^*_{\mathrm{B,p},\mathsf{N}}, \tag{215}$$

where step (a) follows from Proposition 16.

Then we have

$$\begin{aligned}0 &\overset{(a)}{=} \lim_{m\to\infty}\limsup_{n_1\to\infty}\limsup_{k_1\to\infty}\sum_{e=(f_i,f_j)\in\mathcal{E}}\sum_{x_e:\,\gamma_e^{(m)}(x_e)>0}\frac{b_{f_i,e}^{(m,n_1,k_1)}(x_e)+b_{f_j,e}^{(m,n_1,k_1)}(x_e)}{2}\cdot\log\frac{2\gamma_e^{(m)}(x_e)}{b_{f_i,e}^{(m,n_1,k_1)}(x_e)+b_{f_j,e}^{(m,n_1,k_1)}(x_e)} \\ &\overset{(b)}{=} \lim_{m,n_1,k_1\to\infty}\sum_{e=(f_i,f_j)\in\mathcal{E}}\sum_{x_e:\,\gamma_e^{(m)}(x_e)>0}\frac{b_{f_i,e}^{(m,n_1,k_1)}(x_e)+b_{f_j,e}^{(m,n_1,k_1)}(x_e)}{2}\cdot\log\frac{2\gamma_e^{(m)}(x_e)}{b_{f_i,e}^{(m,n_1,k_1)}(x_e)+b_{f_j,e}^{(m,n_1,k_1)}(x_e)},\end{aligned}$$

where step (a) follows from the inequalities in (210) and similar derivations in (214) and (215):

$$F^{\mathrm{alt},*}_{\mathrm{B,d},\mathsf{N}} = \lim_{m,n_1,k_1\to\infty} F^{(1)}_{\mathrm{B,p},\mathsf{N}}(\boldsymbol{b}_{\mathcal{F}}^{(m,n_1,k_1)})$$

$$
\begin{aligned}
&\leq \lim_{m} \hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}) \\
&\quad + \lim_{m} \limsup_{n_1\to\infty} \limsup_{k_1\to\infty} \sum_{e=(f_i,f_j)\in\mathcal{E}} \sum_{x_e:\,\gamma_e^{(m)}(x_e)>0} \frac{b^{(m,n_1,k_1)}_{f_i,e}(x_e)+b^{(m,n_1,k_1)}_{f_j,e}(x_e)}{2} \log \frac{2\gamma_e^{(m)}(x_e)}{b^{(m,n_1,k_1)}_{f_i,e}(x_e)+b^{(m,n_1,k_1)}_{f_j,e}(x_e)} \\
&\leq \lim_{m\to\infty} \hat{F}^{\mathrm{alt}}_{\mathrm{B,d,N}}(\boldsymbol{\gamma}^{(m)}) \\
&= F^{\mathrm{alt},*}_{\mathrm{B,d,N}},
\end{aligned}
$$

and where step (b) follows from the fact that the subsequence $\{b_f^{(m,n_1,k_1)}(\boldsymbol{x}_f)\}_{n_1,\,k_1\in\mathbb{Z}_{>0}}$ converges in $\mathcal{B}_{\mathcal{F}}(\mathsf{N})$ for all $\boldsymbol{x}_f,\ f\in\mathcal{F}$, and the K-L divergence is continuous. By Pinsker's inequality (see, e.g., [20, Theorem 2.33]), we have

$$
\begin{aligned}
0 &= \lim_{m,n_1,k_1\to\infty} \sum_{e=(f_i,f_j)\in\mathcal{E}} \sum_{x_e:\,\gamma_e^{(m)}(x_e)>0} \frac{b^{(m,n_1,k_1)}_{f_i,e}(x_e)+b^{(m,n_1,k_1)}_{f_j,e}(x_e)}{2} \cdot \log \frac{b^{(m,n_1,k_1)}_{f_i,e}(x_e)+b^{(m,n_1,k_1)}_{f_j,e}(x_e)}{2\gamma_e^{(m)}(x_e)} \\
&\geq \frac{1}{2} \lim_{m,n_1,k_1\to\infty} \sum_{e=(f_i,f_j)\in\mathcal{E}} \left(\sum_{x_e:\,\gamma_e^{(m)}(x_e)>0} \left| \frac{b^{(m,n_1,k_1)}_{f_i,e}(x_e)+b^{(m,n_1,k_1)}_{f_j,e}(x_e)}{2} - \gamma_e^{(m)}(x_e) \right| \right)^2 \\
&\geq 0.
\end{aligned} \tag{216}
$$

By the definition of $\boldsymbol{\gamma}^{(m)}$ in Item 2 in Definition 17, the sequence $\{\boldsymbol{\gamma}^{(m)}\}_m$ is bounded. Thus there exists a subsequence of $\{\boldsymbol{\gamma}^{(m)}\}_m$, denoted by $\{\boldsymbol{\gamma}^{(m_1)}\}_{m_1}$, such that the following limits exist

$$
\lim_{m_1\to\infty} \gamma_e^{(m_1)}(x_e), \qquad x_e\in\mathcal{X}_e,\ e\in\mathcal{E},
$$

$$
\lim_{m_1,n_1,k_1\to\infty} b_{f_i}^{(m_1,n_1,k_1)}(\boldsymbol{x}_f), \qquad \boldsymbol{x}_f\in\mathcal{X}_f,\ f\in\mathcal{E}.
$$

By the equalities in (212) and the inequalities in (216), we have

$$
\lim_{m_1\to\infty} \gamma_e^{(m_1)}(x_e) = \lim_{m_1,n_1,k_1\to\infty} b^{(m_1,n_1,k_1)}_{f_i,e}(x_e) = \lim_{m_1,n_1,k_1\to\infty} b^{(m_1,n_1,k_1)}_{f_j,e}(x_e), \qquad x_e\in\mathcal{X}_e,\ e=(f_i,f_j)\in\mathcal{E},
$$

which implies

$$
\left(\lim_{m_1,n_1,k_1\to\infty} \boldsymbol{b}_{\mathcal{F}}^{(m_1,n_1,k_1)} \right) \in \mathcal{B}_{\mathcal{F}}(\mathsf{N}).
$$

By (215), we have

$$
\lim_{m_1,n_1,k_1\to\infty} F^{(1)}_{\mathrm{B,p,N}}(\boldsymbol{b}_{\mathcal{F}}^{(m_1,n_1,k_1)}) = F^{\mathrm{alt},*}_{\mathrm{B,d,N}} = F^{*}_{\mathrm{B,d,N}} = F^{*}_{\mathrm{B,p,N}}.
$$

APPENDIX S
PROOF OF THEOREM 61

In this proof, we only consider the variable $x_e\in\mathcal{S}_e^c$, i.e.,

$$
\gamma_e^*(x_e)\in\mathbb{R}_{>0}, \qquad \forall x_e\in\mathcal{S}_e^c,
$$

where $\mathcal{S}_e^c$ and $\boldsymbol{\gamma}^*$ are defined in (136) and (135), respectively. The proof of (143) and (144) can be obtained by Theorem 57 and the fact that $\left\{\boldsymbol{b}_{\mathcal{F}}^{(m_2,n_2,k_2)}\right\}_{m_2,n_2,k_2}$ is a subsequence of $\left\{\boldsymbol{b}_{\mathcal{F}}^{(m_1,n_1,k_1)}\right\}_{m_2,n_2,k_2}$. In the remaining part of the proof, we focus on proving (141) and (142).

By Theorem 57, the subsequence (m_1,n_1,k_1) has the following property: for any small enough $\epsilon,c\in\mathbb{R}_{>0}$, such that

$$
0<\epsilon+c<\sqrt{\gamma_e^*(x_e)}, \qquad x_e\in\mathcal{S}_e^c,\ e\in\mathcal{E}, \tag{217}
$$

there exist integers M_1, $N_1(M_1)$, and $K_1(M_1,N_1)$ satisfying

$$
\left|\sqrt{\gamma_e^{(m_1)}(x_e)}-\sqrt{\gamma_e^*(x_e)}\right|\leq c, \qquad x_e\in\mathcal{S}_e^c,\ e\in\mathcal{E}, \qquad m_1\geq M_1, \tag{218}
$$

$$\left| \frac{b_{f_i,e}^{(m_1,n_1,k_1)}(x_e)}{\sqrt{\gamma_e^{(m_1)}(x_e)}} - \sqrt{\gamma_e^{(m_1)}(x_e)} \right| \leq \epsilon, \qquad x_e \in \mathcal{S}_e^c,\, e \in \mathcal{E}, \qquad m_1 \geq M_1,\, n_1 \geq N_1(M_1),\, k_1 \geq K_1(M_1, N_1). \tag{219}$$

This implies that for all $x_e \in \mathcal{S}_e^c$, $e \in \mathcal{E}$, $m_1 \geq M_1$, $n_1 \geq N_1(M_1)$, $k_1 \geq K_1(N_1, M_1)$, it holds that

$$\sqrt{\gamma_e^{(m_1)}(x_e)} > \epsilon,$$
$$0 < \sqrt{\gamma_e^{(m_1)}(x_e)} - \epsilon \leq \frac{b_{f_i,e}^{(m_1,n_1,k_1)}(x_e)}{\sqrt{\gamma_e^{(m_1)}(x_e)}} \leq \sqrt{\gamma_e^{(m_1)}(x_e)} + \epsilon.$$

By the definition of $\mu_{e\to f}^{(m_1,n_1,k_1)}$ in (137) and the above inequalities, we have

$$\frac{\sqrt{\gamma_e^{(m_1)}(x_e)} - \epsilon}{\exp\Big(\alpha^{(k_1)} \cdot \lambda_{e,f_i}^{(n_1)}(x_e)\Big)} \leq \frac{b_{f_i,e}^{(m_1,n_1,k_1)}(x_e)}{\mu_{e\to f_i}^{(m_1,n_1,k_1)}(x_e)} \leq \frac{\sqrt{\gamma_e^{(m_1)}(x_e)} + \epsilon}{\exp\Big(\alpha^{(k_1)} \cdot \lambda_{e,f_i}^{(n_1)}(x_e)\Big)},$$
$$\frac{\sqrt{\gamma_e^{(m_1)}(x_e)} - \epsilon}{\sqrt{\gamma_e^{(m_1)}(x_e)}} \cdot \mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e) \leq \frac{b_{f_i,e}^{(m_1,n_1,k_1)}(x_e)}{\mu_{e\to f_i}^{(m_1,n_1,k_1)}(x_e)} \leq \frac{\sqrt{\gamma_e^{(m_1)}(x_e)} + \epsilon}{\sqrt{\gamma_e^{(m_1)}(x_e)}} \cdot \mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e), \tag{220}$$
$$\sum_{x_e\in\mathcal{S}_e^c} \frac{\sqrt{\gamma_e^{(m_1)}(x_e)} - \epsilon}{\sqrt{\gamma_e^{(m_1)}(x_e)}} \cdot \mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e) \leq \sum_{x_e\in\mathcal{S}_e^c} \frac{b_{f_i,e}^{(m_1,n_1,k_1)}(x_e)}{\mu_{e\to f_i}^{(m_1,n_1,k_1)}(x_e)} \leq \sum_{x_e\in\mathcal{S}_e^c} \frac{\sqrt{\gamma_e^{(m_1)}(x_e)} + \epsilon}{\sqrt{\gamma_e^{(m_1)}(x_e)}} \cdot \mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e),$$
$$\Big(1 - \max_{z_e\in\mathcal{S}_e^c} \frac{\epsilon}{\gamma_e^{(m_1)}(z_e)}\Big) \cdot Z_{\mu_{e\to f_j}}^{(m_1,n_1,k_1)} \leq \sum_{x_e\in\mathcal{S}_e^c} \frac{b_{f_i,e}^{(m_1,n_1,k_1)}(x_e)}{\mu_{e\to f_i}^{(m_1,n_1,k_1)}(x_e)} \overset{(a)}{=} C_{e\to f_j}^{(m_1,n_1,k_1)} \leq \Big(1 + \max_{z_e\in\mathcal{S}_e^c} \frac{\epsilon}{\gamma_e^{(m_1)}(z_e)}\Big) \cdot Z_{\mu_{e\to f_j}}^{(m_1,n_1,k_1)}, \tag{221}$$

where the inequalities in (220) follow from the definition of $\boldsymbol{\lambda}_{e,f}$ in (15) and the definition of $\mu_{e\to f}^{(m,n,k)}$, where the inequalities in (221) follow from the definition of $Z_{\mu_{e\to f}}^{(m,n,k)}$ in (138), and where step (a) follows from the definition of $C_{e\to f_j}^{(m,n,k)}$ in (140). We define

$$\epsilon' \triangleq \max_{z_e\in\mathcal{S}_e^c,\, e\in\mathcal{E}} \frac{\epsilon}{\gamma_e^*(z_e) - c} = \frac{\epsilon}{\min_{z_e\in\mathcal{S}_e^c,\, e\in\mathcal{E}}\ \gamma_e^*(z_e) - c}. \tag{222}$$

Then by (217) and (218), we have

$$0 < \epsilon' < 1, \qquad \epsilon' \geq \max_{z_e\in\mathcal{S}_e^c,\, e\in\mathcal{E}}\ \frac{\epsilon}{\gamma_e^{(m_1)}(z_e)} = \frac{\epsilon}{\min_{z_e\in\mathcal{S}_e^c,\, e\in\mathcal{E}}\ \gamma_e^{(m_1)}(z_e)}, \qquad m_1 \geq M_1. \tag{223}$$

It holds that

$$\begin{aligned}
\mu_{e\to f_j,\mathrm{SPA}}^{(m_1,n_1,k_1)}(x_e) &\overset{(a)}{=} \frac{1}{C_{e\to f_j}^{(m_1,n_1,k_1)}} \cdot \frac{b_{f_i,e}^{(m_1,n_1,k_1)}(x_e)}{\mu_{e\to f_i}^{(m_1,n_1,k_1)}(x_e)} \\
&\overset{(b)}{\geq} \frac{\sqrt{\gamma_e^{(m_1)}(x_e)} - \epsilon}{\sqrt{\gamma_e^{(m_1)}(x_e)}} \cdot \Big(1 + \max_{z_e\in\mathcal{S}_e^c} \frac{\epsilon}{\gamma_e^{(m_1)}(z_e)}\Big)^{-1} \cdot \frac{\mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e)}{Z_{\mu_{e\to f_j}}^{(m_1,n_1,k_1)}} \\
&\overset{(c)}{\geq} \frac{\sqrt{\gamma_e^{(m_1)}(x_e)} - \epsilon}{\sqrt{\gamma_e^{(m_1)}(x_e)}} \cdot \Big(1 + \epsilon'\Big)^{-1} \cdot \frac{\mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e)}{Z_{\mu_{e\to f_j}}^{(m_1,n_1,k_1)}},
\end{aligned}$$

where step (a) follows from the definition of $\mu_{e\to f_j,\mathrm{SPA}}^{(m_1,n_1,k_1)}$ in (139), where step (b) follows from the inequalities in (220) and (221), and where step (c) follows from (223). Similarly, we obtain

$$\begin{aligned}
\mu_{e\to f_j,\mathrm{SPA}}^{(m_1,n_1,k_1)}(x_e) &\overset{(a)}{\leq} \frac{\sqrt{\gamma_e^{(m_1)}(x_e)} + \epsilon}{\sqrt{\gamma_e^{(m_1)}(x_e)}} \cdot \Big(1 - \max_{z_e\in\mathcal{S}_e^c} \frac{\epsilon}{\gamma_e^{(m_1)}(z_e)}\Big)^{-1} \cdot \frac{\mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e)}{Z_{\mu_{e\to f_j}}^{(m_1,n_1,k_1)}} \\
&\overset{(b)}{\leq} \frac{\sqrt{\gamma_e^{(m_1)}(x_e)} + \epsilon}{\sqrt{\gamma_e^{(m_1)}(x_e)}} \cdot \Big(1 - \epsilon'\Big)^{-1} \cdot \frac{\mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e)}{Z_{\mu_{e\to f_j}}^{(m_1,n_1,k_1)}},
\end{aligned}$$

where step (a) follows from the inequalities in (220) and (221) and where step (b) follows from (223). The above inequalities imply

$$\begin{aligned}&\left(\frac{\sqrt{\gamma_e^{(m_1)}(x_e)}-\epsilon}{\sqrt{\gamma_e^{(m_1)}(x_e)}}\cdot\left(1+\epsilon'\right)^{-1}-1\right)\cdot\frac{\mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e)}{Z_{\mu_{e\to f_j}}^{(m_1,n_1,k_1)}}\\&\leq \mu_{e\to f_i,\mathrm{SPA}}^{(m_1,n_1,k_1)}(x_e)-\frac{\mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e)}{Z_{\mu_{e\to f_j}}^{(m_1,n_1,k_1)}}\\&\leq\left(\frac{\sqrt{\gamma_e^{(m_1)}(x_e)}+\epsilon}{\sqrt{\gamma_e^{(m_1)}(x_e)}}\cdot\left(1-\epsilon'\right)^{-1}-1\right)\cdot\frac{\mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e)}{Z_{\mu_{e\to f_j}}^{(m_1,n_1,k_1)}}.\end{aligned}\tag{224}$$

In particular, we have

$$\begin{aligned}\left(\frac{\sqrt{\gamma_e^{(m_1)}(x_e)}-\epsilon}{\sqrt{\gamma_e^{(m_1)}(x_e)}}\cdot\left(1+\epsilon'\right)^{-1}-1\right)&\geq\left(\left(1-\max_{z_e\in\mathcal{S}_e^c}\frac{\epsilon}{\gamma_e^{(m_1)}(z_e)}\right)\cdot\left(1+\epsilon'\right)^{-1}-1\right)\\&\overset{(a)}{\geq}\frac{1-\epsilon'}{1+\epsilon'}-1\\&=-\frac{2\epsilon'}{1+\epsilon'},\end{aligned}\tag{225}$$

where step (a) follows from the property of ϵ' in (223). Similarly, we obtain

$$\begin{aligned}\left(\frac{\sqrt{\gamma_e^{(m_1)}(x_e)}+\epsilon}{\sqrt{\gamma_e^{(m_1)}(x_e)}}\cdot\left(1-\epsilon'\right)^{-1}-1\right)&\leq\left(\left(1+\max_{z_e\in\mathcal{S}_e^c}\frac{\epsilon}{\gamma_e^{(m_1)}(z_e)}\right)\cdot\left(1-\epsilon'\right)^{-1}-1\right)\\&\overset{(a)}{\leq}\frac{1+\epsilon'}{1-\epsilon'}-1\\&=\frac{2\epsilon'}{1-\epsilon'},\end{aligned}\tag{226}$$

where step (a) again follows from the property of ϵ' in (223). By the fact that $\mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e)\geq 0$ for all x_e as defined in (137), we have

$$0\leq\frac{\mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e)}{Z_{\mu_{e\to f_j}}^{(m_1,n_1,k_1)}}\leq 1,\qquad x_e\in\mathcal{S}_e^c,\ e\in\mathcal{E}.\tag{227}$$

Then by the inequalities in (224)–(227), it holds that

$$-\frac{2\epsilon'}{1+\epsilon'}\leq-\frac{2\epsilon'}{1+\epsilon'}\cdot\frac{\mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e)}{Z_{\mu_{e\to f_j}}^{(m_1,n_1,k_1)}}\leq\mu_{e\to f_j,\mathrm{SPA}}^{(m_1,n_1,k_1)}(x_e)-\frac{\mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e)}{Z_{\mu_{e\to f_j}}^{(m_1,n_1,k_1)}}\leq\frac{2\epsilon'}{1-\epsilon'}\cdot\frac{\mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e)}{Z_{\mu_{e\to f_j}}^{(m_1,n_1,k_1)}}\leq\frac{2\epsilon'}{1-\epsilon'}.\tag{228}$$

For any small $\epsilon''\in\mathbb{R}_{>0}$, because of

$$\sqrt{\gamma_e^*(x_e)}>0,\qquad x_e\in\mathcal{S}_e^c,\ e\in\mathcal{E},$$

we further require that ϵ and c satisfiy

$$0<\epsilon\leq\left(\min_{z_e\in\mathcal{S}_e^c,\,e\in\mathcal{E}}\ \gamma_e^*(z_e)-c\right)\cdot\frac{\epsilon''}{2+\epsilon''}.\tag{229}$$

Then by the definition of ϵ' in (222), we have

$$\epsilon'\leq\frac{\epsilon''}{2+\epsilon''}\overset{(a)}{\Rightarrow}\frac{2\epsilon'}{1-\epsilon'}\leq\epsilon''.\tag{230}$$

where step (a) follows from $0<\epsilon'<1$ as stated in (223). To sum up, we require that ϵ and c satisfy (217) and (229). There exists $(M_1,N_1(M_1),K_1(M_1,N_1))$ satisfying (218) and (219) for these scalars ϵ and c. Because of the inequalities (228) and (230) and

$$\frac{2\epsilon'}{1-\epsilon'}\geq\frac{2\epsilon'}{1+\epsilon'},$$

we know that

$$\left| \mu_{e\to f_j,\mathrm{SPA}}^{(m_1,n_1,k_1)}(x_e) - \frac{\mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e)}{Z_{\mu_{e\to f_j}}^{(m_1,n_1,k_1)}} \right| \leq \epsilon'', \qquad x_e \in \mathcal{S}_e^c,\, e = (f_i, f_j) \in \mathcal{E}, \qquad m_1 \geq M_1,\, n_1 \geq N_1(M_1),\, k_1 \geq K_1(M_1, N_1),$$

which means that

$$\lim_{m_1,n_1,k_1\to\infty} \left(\mu_{e\to f_j,\mathrm{SPA}}^{(m_1,n_1,k_1)}(x_e) - \frac{\mu_{e\to f_j}^{(m_1,n_1,k_1)}(x_e)}{Z_{\mu_{e\to f_j}}^{(m_1,n_1,k_1)}} \right) = 0, \qquad x_e \in \mathcal{S}_e^c,\, e = (f_i, f_j) \in \mathcal{E}. \tag{231}$$

Similarly, we obtain

$$\lim_{m_1,n_1,k_1\to\infty} \left(\mu_{e\to f_i,\mathrm{SPA}}^{(m_1,n_1,k_1)}(x_e) - \frac{\mu_{e\to f_i}^{(m_1,n_1,k_1)}(x_e)}{Z_{\mu_{e\to f_i}}^{(m_1,n_1,k_1)}} \right) = 0, \qquad x_e \in \mathcal{S}_e^c,\, e = (f_i, f_j) \in \mathcal{E}. \tag{232}$$

As defined in (139), the vector $\mu_{e\to f,\mathrm{SPA}}^{(m_1,n_1,k_1)}(x_e)$ satisfies

$$0 \leq \mu_{e\to f,\mathrm{SPA}}^{(m_1,n_1,k_1)}(x_e) \leq 1, \qquad x_e \in \mathcal{S}_e^c,\, e \in \partial f,\, f \in \mathcal{F}, \quad n_1, k_1 \in \mathbb{Z}_{>0},$$

i.e., it is bounded. By the Bolzano–Weierstrass theorem [23, Theorem 3.4.8], we can find a subsequence $\{m_2, n_2, k_2\}$ of $\{m_1, n_1, k_1\}$ such that the subsequences $\left\{\mu_{e\to f,\mathrm{SPA}}^{(m_2,n_2,k_2)}(x_e)\right\}_{m_2,n_2,k_2\in\mathbb{Z}_{>0}}$ and $\left\{\mu_{e\to f}^{(m_2,n_2,k_2)}(x_e)/Z_{\mu_{e\to f}}^{(m_2,n_2,k_2)}\right\}_{m_2,n_2,k_2\in\mathbb{Z}_{>0}}$ converge for all $x_e \in \mathcal{S}_e^c$, $e \in \partial f$, $f \in \mathcal{F}$. Combining with (231) and (232), we have

$$\lim_{m_2,n_2,k_2\to\infty} \mu_{e\to f_j,\mathrm{SPA}}^{(m_2,n_2,k_2)}(x_e) = \lim_{m_2,n_2,k_2\to\infty} \frac{\mu_{e\to f_j}^{(m_2,n_2,k_2)}(x_e)}{Z_{\mu_{e\to f_j}}^{(m_2,n_2,k_2)}}, \qquad x_e \in \mathcal{S}_e^c,\, e = (f_i, f_j) \in \mathcal{E},$$

$$\lim_{m_2,n_2,k_2\to\infty} \mu_{e\to f_i,\mathrm{SPA}}^{(m_2,n_2,k_2)}(x_e) = \lim_{m_2,n_2,k_2\to\infty} \frac{\mu_{e\to f_i}^{(m_2,n_2,k_2)}(x_e)}{Z_{\mu_{e\to f_i}}^{(m_2,n_2,k_2)}} \qquad x_e \in \mathcal{S}_e^c,\, e = (f_i, f_j) \in \mathcal{E}.$$

REFERENCES

[1] F. R. Kschischang, B. J. Frey, and H.-. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

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

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

[4] H. Wymeersch, *Iterative Receiver Design*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[5] M. Mézard and A. Montanari, *Information, Physics and Computation*. Oxford, U.K.: Oxford Univ. Press, 2009.

[6] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[7] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Uncertainty in Artificial Intelligence*, San Francisco, CA, USA, Jul. 1999, p. 467–475.

[8] T. Heskes, "Stable fixed points of loopy belief propagation are local minima of the Bethe free energy," in *Proc. Neural Information Processing Systems (NIPS)*, Vancouver, Canada, Dec. 2003, pp. 359–366.

[9] 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.

[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] H. D. Pfister and P. O. Vontobel, "On the relevance of graph covers and zeta functions for the analysis of SPA decoding of cycle codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 3000–3004.

[12] C. Knoll, D. Mehta, T. Chen, and F. Pernkopf, "Fixed points of belief propagation—an analysis via polynomial homotopy continuation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2124–2136, Sep. 2017.

[13] C. Knoll and P. Franz, "Belief propagation: Accurate marginals or accurate partition function–where is the difference?" in *Uncertainty in Artificial Intelligence*, 2020, pp. 627–636.

[14] J. M. Walsh, P. A. Regalia, and C. R. Johnson, Jr, "Turbo decoding as iterative constrained maximum-likelihood sequence detection," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5426–5437, Dec. 2006.

[15] P. A. Regalia and J. M. Walsh, "Optimality and duality of the turbo decoder," *Proc. IEEE*, vol. 95, no. 6, p. 1362–1377, Jun. 2007.

[16] T. Heskes, "On the uniqueness of loopy belief propagation fixed points," *Neural Comput.*, vol. 16, no. 11, p. 2379–2413, Nov. 2004.

[17] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundation and Trends in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.

[18] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. USA: Cambridge University Press, 2012.

[19] D. P. Bertsekas, *Nonlinear Programming*, 3rd ed. Belmont, MA, USA: Athena Scientific, 2016.

[20] R. W. Yeung, *Information Theory and Network Coding*, 1st ed. Boston, MA: Springer, 2008.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.

[22] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. Baltimore. MD: JHU Press, 2013.

[23] R. G. Bartle and D. Sherbert, *Introduction to Real Analysis*, 4th ed. New York, USA: J. Wiley, 2010.