On the Relationship Between the 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 Enginerring
The Chinese University of Hong Kong
yuwen.huang@ieee.org, pascal.vontobel@ieee.org

ITW 2022
Mumbai, India

Outline

Outline

- **Overview of the main results**

  Standard normal factor graphs (S-NFGs)

  The sum-product algorithm (SPA)

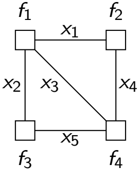
  The primal and dual formulations of the Bethe partition function

  Comparing different dualizations

  Comparison of Yedidia et al.'s results and our results

Overview of standard factor graphs (S-FGs)

- The standard factor graph (S-FG) N consists of
  1. **nonnegative-valued** local functions $f_1, \ldots, f_4$;
  2. edges $1, \ldots, 5$;
  3. alphabets $\mathcal{X}_1, \ldots, \mathcal{X}_5$ for variables $x_1, \ldots, x_5$, respectively.



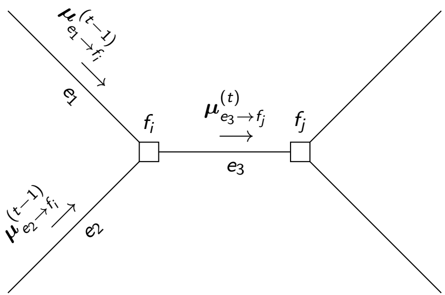
- The global function for N:

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

- We want to approximate the **partition function** of N:

$$Z(\mathrm{N}) \triangleq \sum_{x_1 \in \mathcal{X}_1, \ldots, x_5 \in \mathcal{X}_5} g(x_1, \ldots, x_5).$$

Overview of the sum-product algorithm (SPA)



Let $e_3 = (f_i, f_j) \in \mathcal{E}$. The message $\boldsymbol{\mu}^{(t)}_{e_3 \to f_j}$ is updated based on

$$\mu^{(t)}_{e_3 \to f_j}(x_{e_3}) \propto \sum_{x_{e_1}, x_{e_2}} f_i(x_{e_1}, x_{e_2}, x_{e_3}) \cdot \mu^{(t-1)}_{e_1 \to f_i}(x_{e_1}) \cdot \mu^{(t-1)}_{e_2 \to f_i}(x_{e_2}).$$

Overview of the main results

Prior work by Yedidia *et al.* in [Yedidia et al., 2005]:

1. For standard factor graph (S-FG) with **positive-valued** local functions only, all **local minima** of the Bethe free energy function correspond to **SPA fixed points**.

Our work:

1. By slightly modifying the S-FG with **nonnegative-valued** local functions if necessary, we relate the **global minimum** of the Bethe free energy function to **an SPA fixed point**.

2. The result is mainly based on a **dual** formulation of the Bethe partition function.

Outline

Introduction to S-NFGs

- Many inference problems can be formulated as computing the **marginals** and **partition function** of some multivariate functions.
- S-NFGs are used to represent the **factorizations** of **nonnegative-valued** multivariate functions.
  - The word "normal" means that the variables are arguments of only **one or two** local functions.

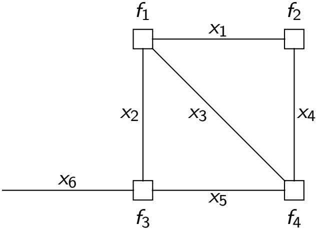
The definition of S-NFGs

The S-NFG $\mathrm{N}(\mathcal{F}, \mathcal{E}, \mathcal{X})$ consists of:

1. the graph $(\mathcal{F}, \mathcal{E})$, where an $f \in \mathcal{F}$ denotes a function node and the associated local function;
2. the alphabet $\mathcal{X} \triangleq \prod_{e \in \mathcal{E}} \mathcal{X}_e$.

An S-NFG consists of two kinds of edges:

1. full edges;
2. half edges.



The definition of S-NFGs

The S-NFG $\mathrm{N}(\mathcal{F}, \mathcal{E}, \mathcal{X})$ consists of:

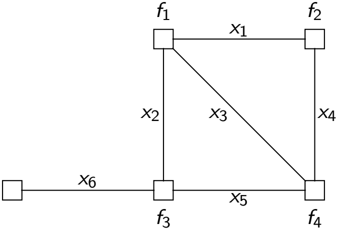
1. the graph $(\mathcal{F}, \mathcal{E})$, where an $f \in \mathcal{F}$ denotes a function node and the associated local function;

2. the alphabet $\mathcal{X} \triangleq \prod_{e \in \mathcal{E}} \mathcal{X}_e$.

An S-NFG consists of two kinds of edges:

1. full edges;

2. half edges.



The definition of S-NFGs

Given $\mathrm{N}(\mathcal{F}, \mathcal{E}, \mathcal{X})$, define

1. the local function: $f : \prod_{e \in \partial f} \mathcal{X}_e \to \mathbb{R}_{\geq 0}$;
2. the global function: $g(\boldsymbol{x}) \triangleq \prod_{f \in \mathcal{F}} f(\boldsymbol{x}_f)$;
3. the partition function: $Z(\mathrm{N}) \triangleq \sum_{\boldsymbol{x}} g(\boldsymbol{x})$.

Outline

Overview of the main results

Standard normal factor graphs (S-NFGs)

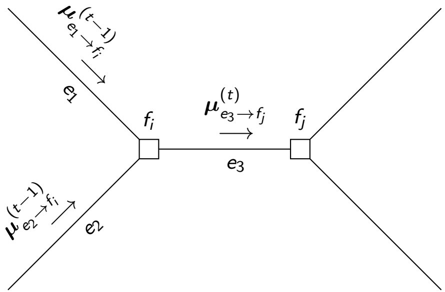
- **The sum-product algorithm (SPA)**

The primal and dual formulations of the Bethe partition function

Comparing different dualizations

Comparison of Yedidia et al.'s results and our results

The SPA



Let t be the iteration index.

1. For $t = 0$, we randomly generate $\boldsymbol{\mu}^{(0)}_{e\to f} \in [0,1]^{|\mathcal{X}_e|} \setminus \{\mathbf{0}\}$.
2. For $t \in \mathbb{Z}_{>0}$ and $e = (f_i, f_j)$, the message from e to f_j is updated according to

$$\mu^{(t)}_{e\to f_j}(x_e) \propto \sum_{\mathbf{z}_{f_i}:\, z_e = x_e} f_i(\mathbf{z}_{f_i}) \cdot \prod_{e' \in \partial f_i \setminus \{e\}} \mu^{(t-1)}_{e'\to f_i}(z_{e'}) \in \mathbb{R}_{\geq 0}.$$

The SPA

For each $e = (f_i, f_j)$, the belief (a.k.a. pseudo-marginal) is defined to be

$$\beta_e^{(t)}(x_e) \triangleq \frac{1}{Z_e(\boldsymbol{\mu}^{(t)})} \cdot \mu_{e \to f_i}^{(t)}(x_e) \cdot \mu_{e \to f_j}^{(t)}(x_e),$$

where the normalization constant Z_e is given by

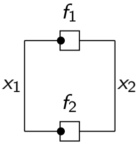
$$Z_e(\boldsymbol{\mu}^{(t)}) \triangleq \sum_{x_e} \mu_{e \to f_i}^{(t)}(x_e) \cdot \mu_{e \to f_j}^{(t)}(x_e).$$

The SPA

- In the case of a **cycle-free** S-NFG, the SPA fixed-point messages provide **exact** marginals and partition function.

- In the case of an S-NFG from **certain** classes of S-NFGs **with cycles**, the SPA fixed-point messages give **good approximations** of the marginals and the partition function.

The SPA on an example S-NFG



We associate the matrices $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ with local functions f_1 and f_2, respectively.

$$\boldsymbol{f}_1 \triangleq \Big(f_1(x_1, x_2)\Big)_{x_1, x_2 \in \mathcal{X}_e} = \begin{pmatrix} f_1(1,1) & \cdots & f_1(1,|\mathcal{X}_2|) \\ \vdots & \ddots & \vdots \\ f_1(|\mathcal{X}_1|,1) & \cdots & f_1(|\mathcal{X}_1|,|\mathcal{X}_2|) \end{pmatrix},$$

$$\boldsymbol{f}_2 \triangleq \Big(f_2(x_1, x_2)\Big)_{x_1, x_2 \in \mathcal{X}_e} = \begin{pmatrix} f_2(1,1) & \cdots & f_2(1,|\mathcal{X}_2|) \\ \vdots & \ddots & \vdots \\ f_2(|\mathcal{X}_1|,1) & \cdots & f_2(|\mathcal{X}_1|,|\mathcal{X}_2|) \end{pmatrix},$$

$$\boldsymbol{M} \triangleq \boldsymbol{f}_1 \cdot \boldsymbol{f}_2^{\mathsf{T}}.$$

The SPA on an example S-NFG

The SPA update rule:

$$\boldsymbol{\mu}_{1\to f_1}^{(t)} \propto \boldsymbol{M} \cdot \boldsymbol{\mu}_{1\to f_1}^{(t-2)}, \qquad \boldsymbol{\mu}_{1\to f_2}^{(t)} \propto \boldsymbol{M}^{\mathsf{T}} \cdot \boldsymbol{\mu}_{1\to f_2}^{(t-2)}.$$

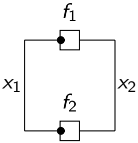
Equivalent to applying the **power method** for the matrix $\boldsymbol{M} \triangleq \boldsymbol{f}_1 \cdot \boldsymbol{f}_2^{\mathsf{T}}$.

At an SPA fixed point $\boldsymbol{\mu}^{(t)}$:

$$\boldsymbol{\mu}_{1\to f_1}^{(t)} \propto \boldsymbol{M} \cdot \boldsymbol{\mu}_{1\to f_1}^{(t)}, \qquad \boldsymbol{\mu}_{1\to f_2}^{(t)} \propto \boldsymbol{M}^{\mathsf{T}} \cdot \boldsymbol{\mu}_{1\to f_2}^{(t)}.$$

The vectors $\boldsymbol{\mu}_{1\to f_1}^{(t)}$ and $\boldsymbol{\mu}_{2\to f_1}^{(t)}$ are the **left and right eigenvectors** of the matrix $\boldsymbol{M}$, respectively.

The SPA on an example S-NFG



Belief on edge 1:

$$\beta_1^{(t)}(x_1) = \frac{1}{Z_1(\boldsymbol{\mu}^{(t)})} \cdot \mu_{1\to f_1}^{(t)}(x_1) \cdot \mu_{1\to f_2}^{(t)}(x_1),$$

where the normalization constant Z_1 is given by

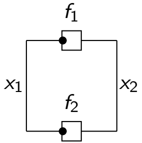
$$Z_1(\boldsymbol{\mu}^{(t)}) = \left(\boldsymbol{\mu}_{1\to f_1}^{(t)}\right)^{\mathsf{T}} \cdot \boldsymbol{\mu}_{1\to f_2}^{(t)}.$$

The SPA on an example S-NFG

Consider specific $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$:

$$\boldsymbol{f}_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{f}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\boldsymbol{M} = \boldsymbol{f}_1 \cdot \boldsymbol{f}_2^{\mathsf{T}} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$



- The largest eigenvalue is **degenerate**.
- The SPA fixed-point messages on edge 1:

$$\boldsymbol{\mu}^{(t)}_{1\to f_1} = (0,\,1)^{\mathsf{T}}, \quad \boldsymbol{\mu}^{(t)}_{1\to f_2} = (1,\,0)^{\mathsf{T}}.$$

- With that, the normalization constant equals

$$Z_1(\boldsymbol{\mu}^{(t)}) = \left(\boldsymbol{\mu}^{(t)}_{1\to f_1}\right)^{\mathsf{T}} \cdot \boldsymbol{\mu}^{(t)}_{1\to f_2} = 0.$$

- This poses **a significant issue** when generalizing the results by Yedidia *et al.* [Yedidia et al., 2005].

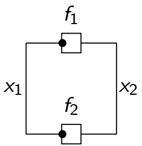
The SPA on an example S-NFG

To address the previous issue, we consider specific $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ such that

$$\boldsymbol{M} = \begin{pmatrix} 1+\delta_2(r) & 1 \\ \delta_1(r) & 1 \end{pmatrix},$$

$$r > 0, \quad \delta_1(r) > 0, \quad \delta_2(r) > 0,$$

$$\lim_{r \downarrow 0} \delta_1(r) = \lim_{r \downarrow 0} \delta_2(r) = 0.$$



- **Perron–Frobenius theory** can be used to show that at the SPA fixed point,

$$\beta_1(x_1) > 0, \quad \forall x_1, \qquad Z_1(\boldsymbol{\mu}^{(t)}) > 0.$$

- Set $r \to 0$. **Different** $\delta_1(r)/\delta_2(r)$ results in **different** SPA fixed-point messages and **different** beliefs $\beta_1(x_1)$.

Outline

The primal formulation

Given $\mathsf{N}(\mathcal{F}, \mathcal{E}, \mathcal{X})$, the **local marginal polytope (LMP)** $\mathcal{B}(\mathsf{N})$ is a collection of vectors

$$\boldsymbol{\beta} \triangleq \left(\{\boldsymbol{\beta}_e\}_{e\in\mathcal{E}}, \{\boldsymbol{\beta}_f\}_{f\in\mathcal{F}} \right)$$

satisfying

1. for $f \in \mathcal{F}$, $\sum_{\boldsymbol{x}_f} \beta_f(\boldsymbol{x}_f) = 1$ **(normalization)**;
2. for $f \in \mathcal{F}$, $\beta_f(\boldsymbol{x}_f) \in \mathbb{R}_{\geq 0}$ **(nonnegativity)**;
3. for $e = (f_i, f_j)$, $\sum_{\boldsymbol{x}_{f_i} :\, x_e = z_e} \beta_{f_i}(\boldsymbol{x}_{f_i}) = \beta_e(z_e) = \sum_{\boldsymbol{x}_{f_j} :\, x_e = z_e} \beta_{f_j}(\boldsymbol{x}_{f_j})$ **(local consistency)**.

$\boldsymbol{\beta} \in \mathcal{B}(\mathsf{N})$ is called a collection of **beliefs (a.k.a. pseudo-marginals)**.

The primal formulation

The **Bethe free energy function** is defined to be

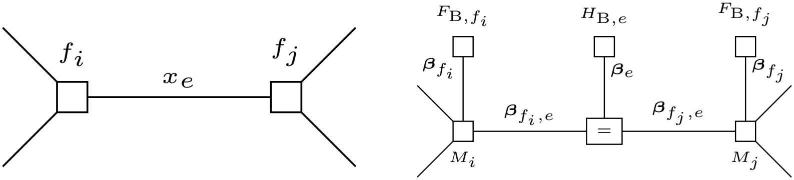
$$F_{\mathrm{B,p,N}} : \ \mathcal{B}(\mathsf{N}) \to \mathbb{R}$$

$$\boldsymbol{\beta} \mapsto \sum_{f} \underbrace{\sum_{\boldsymbol{x}_f} \beta_f(\boldsymbol{x}_f) \cdot \log \frac{\beta_f(\boldsymbol{x}_f)}{f(\boldsymbol{x}_f)}}_{F_{\mathrm{B},f}(\boldsymbol{\beta}_f)} - \sum_{e} \underbrace{\sum_{x_e} \beta_e(x_e) \cdot \log \beta_e(x_e)}_{H_{\mathrm{B},e}(\boldsymbol{\beta}_e)} .$$

The **Bethe approximation of the partition function** $Z(\mathsf{N})$, called the Bethe partition function, is defined to be

$$Z^*_{\mathrm{B,p,N}} \triangleq \exp\left(- \min_{\boldsymbol{\beta}} F_{\mathrm{B,p,N}}(\boldsymbol{\beta}) \right).$$

Factor graphs of the primal formulation



- LHS: part of an S-NFG of interest.
- RHS: part of an NFG whose global function is equal to the **Bethe free energy** function.
  - The global function of this NFG equals the **sum (not the product)** of the local functions.

The primal formulation

When the S-NFG N is **cycle-free**,

1. the function $F_{\mathrm{B,p,N}}(\boldsymbol{\beta})$ is **convex** [Heskes, 2004, Corollary 1];

2. the Bethe partition function $Z^*_{\mathrm{B,p,N}}$ satisfies

$$Z^*_{\mathrm{B,p,N}} = \exp\left(- \min_{\boldsymbol{\beta}} F_{\mathrm{B,p,N}}(\boldsymbol{\beta}) \right) = Z(\mathsf{N});$$

3. the elements in the collection of beliefs

$$\boldsymbol{\beta}^* \in \operatorname{argmin} F_{\mathrm{B,p,N}}(\boldsymbol{\beta})$$

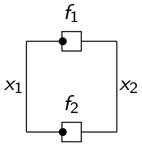
   equal the marginals induced by N [Yedidia et al., 2005, Proposition 3].

The primal formulation

Consider specific $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ associated with function nodes f_1 and f_2:

$$\boldsymbol{f}_1 = \begin{pmatrix} 1 & 1 \\ \delta_1(r) & 1 \end{pmatrix}, \; \boldsymbol{f}_2 = \begin{pmatrix} 1 & \delta_2(r) \\ \delta_3(r) & 1 \end{pmatrix},$$

$$r > 0, \quad \delta_1(r) > 0, \quad \delta_1(r) > 0, \quad \delta_3(r) > 0,$$



$$\lim_{r \downarrow 0} \delta_1(r) = \lim_{r \downarrow 0} \delta_2(r) = \lim_{r \downarrow 0} \delta_3(r) = 0.$$

---

1. Apply [Yedidia et al., 2005, Theorem 3] to this modified S-NFG.
2. Let $r \to 0$.
3. Relate **the global minimum** of the Bethe free energy function to **an SPA fixed point** for the original S-NFG with $\boldsymbol{f}_1 = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $\boldsymbol{f}_2 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

The dual formulation

A dual formulation of the Bethe partition function was proposed in [Yedidia et al., 2005, Walsh et al., 2006, Regalia and Walsh, 2007].

Another dual formulation was presented in [Heskes, 2003, Section 4]:

$$Z^{*}_{\mathrm{B,p,N}} = \max \min \ldots$$

- The dual formulation in [Heskes, 2003, Section 4] is not **well defined**. Heskes did not analyze the optimal values' locations.
- Our contribution is to introduce a **well-defined problem** and study the optimal value's locations in [Huang and Vontobel, 2022, Section III].

The definition of the dual formulation

1. For every edge $e = (f_i, f_j) \in \mathcal{E}$,

$$\boldsymbol{\lambda}_e = \Big(\lambda_e(x_e)\Big)_{x_e} \in \mathbb{R}^{|\mathcal{X}_e|}, \qquad \boldsymbol{\lambda}_{e,f_i} = \boldsymbol{\lambda}_e,\ \boldsymbol{\lambda}_{e,f_j} = -\boldsymbol{\lambda}_e,$$

$$\boldsymbol{\gamma}_e = \Big(\gamma_e(x_e)\Big)_{x_e} \in \mathbb{R}_{\geq 0}^{|\mathcal{X}_e|}, \qquad \sum_{x_e} \gamma_e(x_e) = 1.$$

2. For every $f \in \mathcal{F}$,

$$Z_f(\boldsymbol{\gamma}_{\partial f}, \boldsymbol{\lambda}_{\partial f}) \triangleq \sum_{\boldsymbol{x}_f} f(\boldsymbol{x}_f) \cdot \prod_{e \in \partial f} \underbrace{\Big(\exp\big(\lambda_{e,f}(x_e)\big) \cdot \sqrt{\gamma_e(x_e)}\Big)}_{\mu_{e\to f}(x_e)}.$$

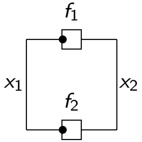
3. For S-NFG N,

$$Z_{\mathrm{B,d,N}}^{\mathrm{alt},*} \triangleq \sup_{\boldsymbol{\gamma}} \inf_{\boldsymbol{\lambda}} \prod_f Z_f(\boldsymbol{\gamma}_{\partial f}, \boldsymbol{\lambda}_{\partial f}).$$

The dual formulation for an example S-NFG

Consider specific $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ associated with function nodes f_1 and f_2:

$$\boldsymbol{f}_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad \boldsymbol{f}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$



There are $\{\gamma^{(m)}\}$ and $\{\boldsymbol{\lambda}^{(n)}\}$ such that

1. $\{\gamma^{(m)}\}$ and $\{\boldsymbol{\lambda}^{(n)}\}$ **converges** to the location of the **optimal** value;
2. an associated message sequence **converges** to a collection of **SPA fixed-point** messages.

Outline

Overview of the main results

Standard normal factor graphs (S-NFGs)

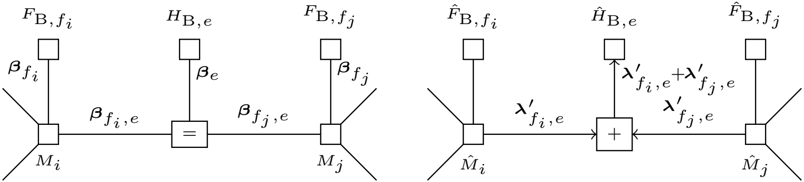
The sum-product algorithm (SPA)

The primal and dual formulations of the Bethe partition function

- **Comparing different dualizations**

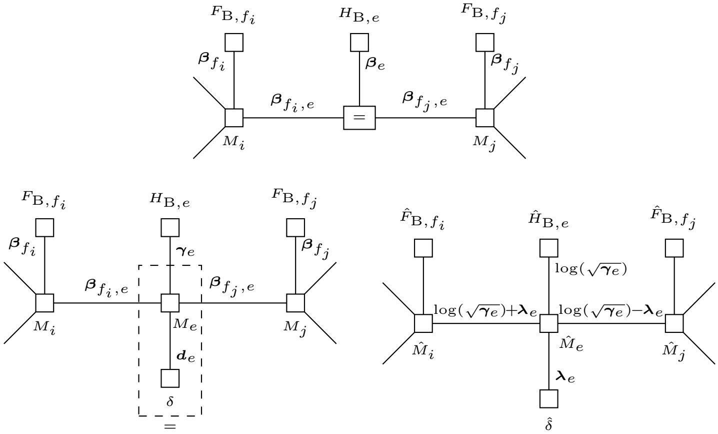
Comparison of Yedidia et al.'s results and our results

The dualization by Yedidia et al.



- **Dualizing** the NFG according to [Yedidia et al., 2005, Walsh et al., 2006, Regalia and Walsh, 2007].
- The details are given in [Yedidia et al., 2005, Section VI] and [Regalia and Walsh, 2007, Section V-C].

The dualization by Heskes



1. **Replacing** the equal-constraint function node.
2. **Dualizing** the resulting NFG.
3. The details are in [Huang and Vontobel, 2022, Appendix C].

Outline

Overview of the main results

Standard normal factor graphs (S-NFGs)

The sum-product algorithm (SPA)

The primal and dual formulations of the Bethe partition function

Comparing different dualizations

Comparison of Yedidia et al.'s results and our results

Comparison of the results

Prior work by Yedidia *et al.* in [Yedidia et al., 2005]:

- For the S-NFG with **positive-valued** local functions only, all **local minima** of the Bethe free energy function correspond to **SPA fixed points** .

Our work:

- By slightly modifying the S-NFG with **nonnegative-valued** local functions if necessary, we relate the **global minimum** of the Bethe free energy function to **an SPA fixed point**.

Selected References I

Heskes, T. (2003).
Stable fixed points of loopy belief propagation are local minima of the Bethe free energy.
In *Proc. Neural Information Processing Systems (NIPS)*, pages 359–366, Vancouver, Canada.

Heskes, T. (2004).
On the uniqueness of loopy belief propagation fixed points.
Neural Comput., 16(11):2379–2413.

Huang, Y. and Vontobel, P. O. (2022).
On the relationship between the global minimum of the Bethe free energy function of a factor graph and sum-product algorithm fixed point (extended version).

Regalia, P. A. and Walsh, J. M. (2007).
Optimality and duality of the turbo decoder.
Proc. IEEE, 95(6):1362–1377.

Selected References II

Walsh, J. M., Regalia, P. A., and Johnson, Jr, C. R. (2006).
Turbo decoding as iterative constrained maximum-likelihood sequence detection.
IEEE Trans. Inf. Theory, 52(12):5426–5437.

Yedidia, J. S., Freeman, W. T., and Weiss, Y. (2005).
Constructing free-energy approximations and generalized belief propagation algorithms.
IEEE Trans. Inf. Theory, 51(7):2282–2312.

Thank you!