Minima and Critical Points of the Bethe Free Energy Are Invariant Under Deformation Retractions of Factor Graphs

Grégoire Sergeant-Perthuis[*1,2] and Léo Boitel[1]

[1]CQSB, Sorbonne Université, Paris, France
[2]Ouragan, Inria Paris, France

Abstract

In graphical models, factor graphs, and more generally energy-based models, the interactions between variables are encoded by a graph, a hypergraph, or, in the most general case, a partially ordered set (poset). Inference on such probabilistic models cannot be performed exactly due to cycles in the underlying structures of interaction. Instead, one resorts to approximate variational inference by optimizing the Bethe free energy. Critical points of the Bethe free energy correspond to fixed points of the associated Belief Propagation algorithm. A full characterization of these critical points for general graphs, hypergraphs, and posets with a finite number of variables is still an open problem. We show that, for hypergraphs and posets with chains of length at most 1, changing the poset of interactions of the probabilistic model to one with the same homotopy type induces a bijection between the critical points of the associated free energy. This result extends and unifies classical results that assume specific forms of collapsibility to prove uniqueness of the critical points of the Bethe free energy.

1 Introduction

Graphical models are probabilistic models that account for conditional independence relations between variables, derived from the connectivity properties of substructures of a graph [10]. The Hammersley-Clifford theorem relates those conditional independence properties to a a way the joint distribution should factor: it is the product of factor, on per clique of the graphs, each factor depending only on variables inside a given clique. Factor graphs are a generalization of graphical models which encode how distribution factor into terms that depend on sub collection of variables. In applications, some of the variables in a factor graph are called *hidden* and others *observed*; computing the posterior of the joint distribution that factorizes according to the graph, conditioned on the observed variables, is computationally intractable when done naively. One instead resorts to an *approximate inference* method, which corresponds to minimizing the Bethe Free Energy under linear constraints. The Belief Propagation algorithm is one algorithm used to find the critical points of the free energy: critical points of the approximate free energy are in correspondence with the fixed points of the algorithm. Therefore, one way to study the critical points of the Bethe free energy is to study the fixed points of the respective

---

[*]gregoire.sergeant-perthuis@sorbonne-universite.fr

Belief Propagation algorithm. These critical points have been observed to be non-unique, and characterizing the set of such points is still ongoing [2–8, 11–13, 16, 20, 26].

The topology of the graph or hypergraph encoding the dependencies of the variables affects the possible set of critical points one can obtain. When the graph underlying the graphical model is acyclic, i.e., possesses no cycles, it is well known that the Bethe free energy has only one critical point, and that approximate inference in this case is exact. In Section 4.2.1 of [22], hypertrees or acyclic hypergraphs are introduced.

In [24], the authors study how pruning the poset can keep the free energy unchanged, and in Section 6.1.1 of [14], a notion of retractable hypergraph is given for which the free energy has only one critical point. Although the operations studied and conditions imposed are of a topological nature, they seem to be dispersed results focused on very particular aspects of the poset. Understanding the impact of the topology of finite factor graphs with cycles on the critical points of the Bethe free energy remains to be explored in a more systematic manner. We aim to address this by characterizing the critical points of this energy in terms of the homotopy type of the hypergraph associated with the factor graph. In this article, we show more generally that, for a good notion of homotopy on factor graphs, the critical points of the Bethe Free Energy are isomorphic.

2 The Bethe Free Energy and the General Belief Propagation algorithms

2.1 Hypergraphs, Factor Graphs, and Posets

Let I be a finite index set of random variables: for each $i \in I$, there is a variable X_i taking values in a finite set E_i. A hypergraph denoted as $\mathscr{H} = (I, A)$ is a finite collection of vertices I, and a finite collection A of subsets $a \subseteq I$ called hyperedges, i.e. A is a subset of the powerset $\mathscr{P}(I)$.

Let us denote by $\mathbb{R}_{\geq 0}$ (respectively $\mathbb{R}_{>0}$) the set of non-negative (respectively strictly positive) real numbers.

Definition 1 (Factor graph [9, 25]). *A factor graph $\mathscr{F} = (\mathscr{H}, E, f)$ is made of a hypergraph $\mathscr{H} = (I, A)$, a collection of finite sets $(E_i; i \in I)$ and a collection of factor $f_a : \prod_{i \in a} E_i \to \mathbb{R}_{\geq 0}$, one for each $a \in A$. In this document, we will call factor graph a couple $\mathscr{F} = (\mathscr{H}, E, H)$, where for each $a \in \mathscr{A}$, $H_a : \prod_{i \in a} E_i \to \mathbb{R}$ is a function that is not necessarily positive; to each H_a we associate the factor $f_a = e^{-H_a}$ which is a strictly positive function. Our convention for factor graphs is more restrictive as we assume that each factor is a strictly positive function, however it is a common assumption [25].*

The standard way to present factor graphs is to introduce a bipartide graph made of variable nodes, $i \in I$, and of factor nodes, $a \in A$, where each factor node is labelled with f_a. For all $i \in I$, and $a \in A$, if $i \in a$ then one adds a directed edge $a \to i$. A hypergraph is a particular example of poset $\mathscr{A}(\mathscr{H})$, which elements are $I \sqcup A$ and relation $i \leq a$ are generated by $i \in a$ with $i \in I$ and $a \in A$. And the Hasse diagram of $\mathscr{A}(\mathscr{H})$ is the directed bipartite graph associated to the factor graph with all arrows reversed, i.e. where every arrow $a \to i$ is replaced by $i \to a$.

2.2 Bethe Free Energy

Let I be a finite set, let $\mathscr{A} \subseteq \mathscr{P}(I)$ be a collection of subsets of I. For each $a \in \mathscr{A}$, consider a factor function $f_a : E_a \to \mathbb{R}_{\geq 0}$, where $E_a = \prod_{i \in a} E_i$ and $E_I = \prod_{i \in I} E_i$. We denote by $\mathbb{P}(E_a)$ the space of probability distributions over E_a; that is, $P \in \mathbb{P}(E_a)$ is a collection $(P(x); x \in E_a)$ such that $\sum_{x \in E_a} P(x) = 1$. Similarly let $P \in \mathbb{P}_{>0}(E_a)$ when $P \in \mathbb{P}(E_a)$ and for all $x \in E_a$, $P(x) > 0$.

A joint distribution $P \in \mathbb{P}(E_I)$ is said to factor according to the collection $\{f_a\}_{a \in \mathscr{A}}$ if, for any $x \in E_I$,

$$P(x) = \prod_{a \in \mathscr{A}} f_a(x_a),$$

where $x_a = (x_i)_{i \in a}$ denotes the restriction of x to the coordinates indexed by a. We denote the projection from $x \in E_a$ to $x_b \in E_b$, with $b \subseteq a$, as π_b^a.

The Möbius function of a poset $\mathscr{A}$ (Proposition 2 in [17]) is the unique collection $\mu = \left(\mu(a,b) \in \mathbb{Z}; b \leq a\right)$ satisfying, for any $a, c \in \mathscr{A}$ with $c \leq a$,

$$\delta_{a,c} = \sum_{b:c \leq b \leq a} \mu(b,c) = \sum_{b:c \leq b \leq a} \mu(a,b).$$

$\delta_{a,c} = 1$ if $a = c$, and 0 otherwise. We will also use the notation $1[C(x)]$ for the indicator function which equals 1 if the condition $C(x)$ is satisfied and 0 otherwise.

The *Generalized Bethe free energy* [15, 25] associated with the pair $(\mathscr{A}, (f_a)_{a \in \mathscr{A}})$ is defined for any collection of probability distributions $Q = (Q_a \in \mathbb{P}(E_a))_{a \in \mathscr{A}}$ as

$$BT_{\mathscr{A},H}(Q) = \sum_{a \in \mathscr{A}} c(a) \left(\mathbb{E}_{Q_a}[H_a] - S(Q_a)\right), \tag{2.1}$$

where for each $a \in \mathscr{A}$,

$$c(a) = \sum_{b \geq a} \mu(b,a), \qquad H_a(x_a) = \sum_{b \subseteq a} -\ln f_b(x_b).$$

We call H_a a Hamiltonian, it is simply a function $H_a : E_a \to \mathbb{R}$. We recover the classical Bethe approximation for graphical models when $\mathscr{A} = \mathscr{A}(G)$ for an undirected graph G (seen as a hypergraph), and for factor graphs when $\mathscr{A} = \mathscr{A}(\mathscr{H})$ for a hypergraph $\mathscr{H}$ (see Chapter 4 [22]). The relevant family of distributions $(Q_a \in \mathbb{P}(E_a))_{a \in \mathscr{A}}$ is the one satisfying marginalization consistency: for all $b \subseteq a$, we require that Q_b is the marginal of Q_a, obtained by summing over the configurations of the variables $X_{a \setminus b} = (X_i \mid i \in a, \ i \notin b)$. This compatibility condition is called in [22] the *polyhedral outer bound* of the marginal polytope, and is denoted by $\mathbb{L}(\mathscr{A})$. More precisely,

$$\mathbb{L}(\mathscr{A}) := \left\{ Q = (Q_a)_{a \in \mathscr{A}} \in \prod_{a \in \mathscr{A}} \mathbb{P}(E_a) \ \middle| \ \forall b \leq a, \ \forall x \in E_b, \ Q_b(x) = \sum_{\substack{y \in E_a \\ y_b = x}} Q_a(y) \right\}. \tag{2.2}$$

The marginal polytope corresponds to all the probability distributions $(Q_a, a \in \mathscr{A})$ that are the marginals of a global $P \in \mathbb{P}(E_I)$. $\mathbb{L}(\mathscr{A})$ is called the polyhedral outer bound, as it

is a polytope defined by linear equalities and inequalities that contains strictly the marginal polytope: there may exist a collection of local distributions $(Q_a)_{a \in \mathscr{A}}$ that do not arise from any global distribution $Q \in \mathbb{P}(E)$. Let us first remark that $\mathbb{L}(\mathscr{A})$ is a compact set, and the Generalized Bethe free energy is continuous on this set. Therefore, it has at least one global minimum.

The interior of $\mathbb{L}(\mathscr{A})$, denoted $\mathbb{L}(\mathscr{A})^{\circ}$, consists of distributions $Q_a \in \mathbb{P}_{>0}(E_a)$, for $a \in \mathscr{A}$, that are compatible by marginalization. Characterizing the critical points of the Generalized Bethe free energy within $\mathbb{L}(\mathscr{A})^{\circ}$ is an important problem in the study of graphical models. In the next section, we will recall and reprove a known result: there is always at least one critical point in this set when the poset $\mathscr{A}$ arises from a factor. The authors found it difficult to present a full and simple proof of this result; that is why we chose to prove the result in the article. We do so by characterizing these points as fixed points of the Generalized Belief Propagation (GBP) algorithm. The proof relies on Brouwer's fixed-point theorem applied to Generalized Belief Propagation (GBP), analogous to its application in the max-product version of Belief Propagation [21]. However, for general posets (of rank, with maximal chains that can be of length greater than 2), this result does not apply. Although GBP maps the interior of $\mathbb{L}(\mathscr{A})$ into itself, it cannot be continuously extended to the boundary, making BP not a continuous operator from $\mathbb{L}(\mathscr{A})$ to itself.

Hamiltonians $(H_a)_{a \in \mathscr{A}}$ and factors $(f_a)_{a \in \mathscr{A}}$ are related by the zeta function of the poset $\mathscr{A}$: $H_a = \sum_{b \subseteq a} -\ln f_b$; this operator has an inverse given by $-\ln f_a = \sum_{b \subseteq a} \mu(b,a) H_b$. Starting from $\ln f$ or H is equivalent; in what follows, we consider Hamiltonians instead of factors.

2.3 The General Belief Propagation algorithm, and existence of critical points of the Bethe Free energy for factor graphs

2.4 The general setting

The General Belief Propagation algorithm is used to find the critical points of the Bethe Free Energy as it is a well know fact that the critical point of the Bethe Free Energy are in correspondance with fix points of the General Belief Propagation algorithm [15, 25]; for a detail proof of this result see Proposition 3 [18].

Let I be a finite set that serves as index for variables $(X_i \in E_i; i \in I)$, each of which takes values in a finite set E_i; let $\mathscr{A} \subseteq \mathscr{P}(I)$ be a collection of subsets of I and denote F the associated graphical presheaf. Let $(H_a : E_a \to \mathbb{R}, a \in \mathscr{A})$ be a collection of Hamiltonians. Let us denote the update rule of the General Belief Propagation algorithm as BP. BP acts on messages that we will now define. In the classical presentation of the algorithm, there are two types of messages at each time $t \in \mathbb{N}^*$. For elements $a, b \in \mathscr{A}$ such that $b \subseteq a$, we have top-down messages $m_{a \to b} \in \mathbb{R}_{>0}^{E_b}$ and bottom-up messages $n_{b \to a} \in \mathbb{R}_{>0}^{E_b}$.

These messages are related as follows:

$$\forall a, b \in \mathscr{A}, \text{s.t. } b \subseteq a, \quad n_{b \to a}^t = \prod_{\substack{c: b \subseteq c \\ c \nsubseteq a}} m_{c \to b}^t \tag{2.3}$$

Beliefs, which are interpreted as probability distributions up to a multiplicative constant are defined as follows:

4

$$\forall a \in \mathscr{A}, \forall x_a \in E_a \quad b_a^t(x_a) \propto e^{-H_a(x_a)} \prod_{\substack{b \in \mathscr{A}: \\ b \subseteq a}} n_{b \to a}^t(x_b) \tag{2.4}$$

where $\propto$ stands for proportional to. The multiplication of function $n_{b \to a}$ that have different domains is made possible because there is an the embedding of $\mathbb{R}^{E_b}$ into $\mathbb{R}^{E_a}$ implicitly implied in the last equation; indeed, for $x \in E_a$ and $f \in \mathbb{R}^{E_b}$, $f : x \mapsto f(x_b)$ defines a function from E_a to $\mathbb{R}$.

For simplicity, we require that b_a be a probability distribution and normalize it accordingly. The update rule is given by,

$$\forall a,b \in \mathscr{A} \text{ s.t. } b \subseteq a, \forall x \in E_b, \quad m_{a \to b}^{t+1}(x) = m_{a \to b}^t(x) \frac{\sum_{y \in E_a : y_b = x} b_a^t(y)}{b_b^t(x)} \tag{2.5}$$

One observes that in the previous Equation 2.5, any normalization of beliefs does not change the update rule.

The update rule can be rewritten in a more condensed manner, updating only the top-down messages, for all $a, b \in \mathscr{A}$, such that $b \leq a$,

$$m_{a \to b}^{t+1}(x) = m_{a \to b}^t(x) \frac{\sum_{y \in E_a : y_b = x} e^{-H_a(y)} \prod_{\substack{c \in \mathscr{A}: \\ c \subseteq a}} \prod_{\substack{d : c \subseteq d \\ d \not\subseteq a}} m_{d \to c}^t(y_c)}{e^{-H_b(x)} \prod_{\substack{c \in \mathscr{A}: \\ c \subseteq b}} \prod_{\substack{d : c \subseteq d \\ d \not\subseteq b}} m_{d \to c}^t(x_c)} \tag{2.6}$$

We denote the collection $(m_{a \to b}; a, b \in \mathscr{A}, b \leq a)$ as m. We denote BP : $\prod_{a,b:b \subseteq a} \mathbb{R}^{E_b} \to \prod_{a,b:b \subseteq a} \mathbb{R}^{E_b}$ as the operator underlying the update rule of Equation 2.6, i.e., we define $\mathrm{BP}(m^t) = m^{t+1}$.

Consider the collection $(C_{a \to b} m_{a \to b}; a, b \in \mathscr{A}, b \leq a)$, where $C_{a \to b}$ is a strictly positive constant, i.e., it does not depend on $x \in E_b$. Then, there is a collection of constants $(C'_{a \to b} > 0; a, b \in \mathscr{A}, b \leq a)$ such that

$$\mathrm{BP}(C_{a \to b} m_{a \to b}; b \leq a) = (C'_{a \to b} \cdot \mathrm{BP}(m)_{a \to b}; b \leq a).$$

Furthermore, the associated beliefs defined by Equation 2.4 remain unchanged under multiplication of $m_{a \to b}$ by a constant $C_{a \to b}$ for all $a, b \in \mathscr{A}$ such that $b \leq a$.

Therefore, BP is an algorithm that preserves the equivalence classes $\{C \cdot m\}$, i.e., it is defined by the relation $m \sim m'$ whenever there is a collection of scalars $(C_{a \to b} \neq 0; a, b \in \mathscr{A}, b \leq a)$ such that

$$m_{a \to b} = C_{a \to b} m'_{a \to b} \quad \text{for any } a, b \in \mathscr{A} \text{ with } b \leq a.$$

We shall denote the equivalence class of m as $[m]$. The action of BP on the equivalence classes of messages is denoted by [BP] and defined as $[\mathrm{BP}]([m]) = [\mathrm{BP}(m)]$.

2.5 The special case of factor graphs

The update rule in Equation 2.6 simplifies in the context of factor graphs, as we will explain now. Let $\mathscr{F}$ be a factor graph and $\mathscr{H} = (I, A)$ the associated hypergraph. We denote by $\mathscr{N}(a)$ the set of vertices $i \in I$ such that $i \in a$, and $\mathscr{N}(i)$ the set of hyperedges a such that

$i \in a$. Note that any $m_{a \to a}^t$ are kept unchanged over time and do not appear in the update rule; one can choose them to be equal to 1, or simply omit them. The only message we must consider are those on arrows $a \in A \to i \in I$; remark that the part of the numerator of the update rule can be written as:

$$\prod_{\substack{c \in \mathscr{A} \\ c \subseteq a}} \prod_{\substack{d:c \subseteq d \\ d \not\subseteq a}} m_{d \to c}^t(x_c) = \prod_{j \in \mathscr{N}(a)} \prod_{b \in \mathscr{N}(j) \backslash a} m_{b \to j}^t(x_j) \tag{2.7}$$

$$= \prod_{b \in \mathscr{N}(i) \backslash a} m_{b \to i}^t(x_i) \cdot \prod_{j \in \mathscr{N}(a) \backslash i} \prod_{b \in \mathscr{N}(j) \backslash a} m_{b \to j}^t(x_j) \tag{2.8}$$

$\mathscr{N}(j) \backslash a$ denotes the set $\mathscr{N}(j)$ without the element a; this is an abuse of notation for $\mathscr{N}(j) \backslash \{a\}$. Similarly for $\mathscr{N}(a) \backslash j$.

Part of the denominator can be rewritten as:

$$\prod_{\substack{c \in \mathscr{A} \\ c \subseteq i}} \prod_{\substack{d:c \subseteq d \\ d \not\subseteq i}} m_{d \to c}^t(x_c) = \prod_{b \in \mathscr{N}(i)} m_{b \to i}^t(x_i) \tag{2.9}$$

Remark that the numerator simplifies as,

$$m_{a \to i}^t(x_i) \prod_{b \in \mathscr{N}(i) \backslash a} m_{b \to i}^t(x_i) = \prod_{b \in \mathscr{N}(i)} m_{b \to i}^t(x_i) \tag{2.10}$$

and this previous term only depends on x_i and therefore factors through the sum in the numerator. In the end the update rule of BP for factor graphs is simply,

$$m_{a \to i}^{t+1}(x_b) = e^{H_i(x_i)} \sum_{y_a: \pi_b^a(y_a) = x_b} e^{-H_a(y_a)} \prod_{j \in \mathscr{N}(a) \backslash i} \prod_{b \in \mathscr{N}(j) \backslash a} m_{b \to j}^t(x_j) \tag{2.11}$$

Let us now remark that if all the messages $m_{a \to i}^t > 0$ for $a \in A$ and $i \in a$, then the messages $m_{a \to i}^{t+1} > 0$ are also strictly positive at time $t + 1$. In particular, the associated beliefs are given by

$$\forall a \in A, \ \forall x_a \in E_a, \quad b_a^t(x_a) \propto e^{-H_a(x_a)} \prod_{i \in a} \prod_{\substack{b:b \neq a \\ i \in b}} m_{b \to i}^t(x_i). \tag{2.12}$$

and

$$\forall a \in I, \ \forall x_i \in E_i, \quad b_i^t(x_i) \propto e^{-H_i(x_i)} \tag{2.13}$$

are strictly positive, i.e., for all $a \in \mathscr{A}$ and $x_a \in E_a$, we have $b_a(x_a) > 0$. For strictly positive messages, the update rule in Equation 2.11 can be rewritten using $\ln m_{a \to i}^t$ denoted as $M_{a \to i}$, the associated update rule is,

$$M_{a \to i}^{t+1}(x_i) = H_i(x_i) + \ln \sum_{y_a: \pi_i^a(y_a) = x_i} \exp \left(-H_a(y_a) + \sum_{j \in \mathscr{N}(a) \backslash i} \sum_{b \in \mathscr{N}(j) \backslash a} M_{b \to j}^t(x_j) \right). \tag{2.14}$$

We denote the update rule of the previous equation, Equation (2.14), as $\log \mathrm{BP}$, i.e.,

$$M_{a\to i}^{t+1}(x_i) = \big(\log \mathrm{BP}(M)\big)_{a\to i}(x_i).$$

As discussed previously, BP is defined on messages up to a multiplicative constant for each message. Similarly, $\log \mathrm{BP}$ is defined up to an additive constant. We denote the associated equivalence class by $[\,\cdot\,]$. Let us recall the notation: for any $a \in A$ and $i \in I$ such that $i \in a$,

$$C_{a\to i}(M) = \sup_{x_i \in E_i} \log \mathrm{BP}_{a\to i}(M).$$

In particular,

$$\log \mathrm{BP}(M - C(M)) = \log \mathrm{BP}(M) + C(M), \qquad [\log \mathrm{BP}(M)] = [\log \mathrm{BP}(M) - C(M)].$$

Note that, by definition of C, for any $a \in A$ and $i \in I$ such that $i \in a$,

$$\sup_{x_i \in E_i} \Big(\log \mathrm{BP}(M)_{a\to i} - C(M)_{a\to i} \Big) = 0.$$

Proposition 1. *The Belief propagation algorithm up to multiplicative constants has at least one fix point such that all messages are strictly positive, i.e. there is a collection $\big(M_{a\to i}^* \in \mathbb{R}^{E_i}; a \in A, i \in I, \ s.t. \ i \in a\big)$ such that,*

$$[M^*] = \log BP([M^*]) \tag{2.15}$$

Proof. The proof of the existence of fixed points of the max-product algorithm on graphs [21] can be adapted into a proof of the existence of fixed points of the Belief Propagation algorithm on factor graphs.

One remarks that,

$$x \in E_a \ \mapsto \ \sum_{j \in \mathcal{N}(a)\setminus i} \ \sum_{b \in \mathcal{N}(j)\setminus a} m_{b\to j}^t(x_j)$$

does not depend on x_i. Let us denote this terms as $g(x_{a\setminus i})$. Therefore, for any two $x, z \in E_i$,

$$\log \mathrm{BP}(M)_{a\to i}(x) = H_i(z) + \big(H_i(x) - H_i(z)\big) + \tag{2.16}$$

$$\ln \sum_{\substack{y \in E_a \\ y_i = x_i}} \exp \big(-H_a(y_{a\setminus i}, z) + \big(H_a(y_{a\setminus i}, x) - H_a(y_{a\setminus i}, z)\big) + g(x_{a\setminus i}) \big).$$

$$\tag{2.17}$$

Therefore,

$$\log \mathrm{BP}(M)_{a\to i}(x) \le \log \mathrm{BP}(M)_{a\to i}(z) + \sup_{x,z \in E_i} |H_i(x) - H_i(z)| + \sup_{y,y_1 \in E_a} |H_a(y) - H_a(y_1)| \tag{2.18}$$

By exchanging x and z, one obtains that

$$\sup_{x,z \in E_i} \big| \log \mathrm{BP}(M)_{a \to i}(x) - \log \mathrm{BP}(M)_{a \to i}(z) \big| \leq \sup_{x,z \in E_i} |H_i(x) - H_i(z)| + \sup_{y,y_1 \in E_a} |H_a(y) - H_a(y_1)|$$
(2.19)

Therefore there is Λ such that

$$\sup_{\substack{a \in A,\, i \in I \\ i \in a}} \sup_{x,y \in E_i} \big| \log \mathrm{BP}(M)_{a \to i}(x) - \log \mathrm{BP}(M)_{a \to i}(y) \big| \leq \Lambda.$$
(2.20)

In particular, one can deduce from Equation 2.20 and the fact that

$$\sup_{x_i \in E_i} \Big(\log \mathrm{BP}(M)_{a \to i}(x) - C(M)_{a \to i} \Big) = 0,$$

that,

$$\sup_{\substack{a \in A \\ i \in a}} \sup_{x \in E_i} \big| \log \mathrm{BP}(M)_{a \to i}(x) - C(M)_{a \to i} \big| \leq \Lambda.$$
(2.21)

Denote by K the subset of messages that are bounded by Λ. In other words,

$$K := \left\{ M \;\middle|\; \sup_{\substack{a \in A \\ i \in a}} \sup_{x \in E_i} \big| M_{a \to i}(x) \big| \leq \Lambda \right\}.$$
(2.22)

And denote by $[K]$ the set of equivalence classes,

$$[K] := \{ [M] \mid M \in K \}.$$
(2.23)

Then, $[\log \mathrm{BP}]$ maps $[K]$ into $[K]$. Moreover, $[K]$ is a convex and compact set. Therefore, by Brouwer's fixed-point theorem, $[\log \mathrm{BP}]$ admits at least one fixed point, denoted by $[M^*]$. $\square$

Corollary 1. *Let $E = \prod_{i \in I} E_i$ and let $\mathscr{H}$ be a factor graph. Let $\mathscr{A}$ be the associated poset of $\mathscr{H}$. Then the generalized Bethe free energy $BT_{\mathscr{A}, \mathscr{H}}$ admits at least one critical point in the interior of the polyhedral outer bound of the marginal polytope $\mathbb{L}(F)$.*

Proof. Let $\mathscr{H} = (I, A)$ be a hypergraph with nodes I and hyperedges A. By Proposition 1, there exists a fixed point $[M^*]$ of $[\log \mathrm{BP}]$. For any $a \in A$ and $i \in I$ such that $i \in a$, denote

$$m^*_{a \to i} \propto e^{M^*_{a \to i}}.$$

Let b be the associated belief, defiend as,

$$\forall a \in A, \ \forall x_a \in E_a, \quad b_a(x_a) \propto e^{-H_a(x_a)} \prod_{i \in a} \prod_{\substack{b \in A \\ b \neq a,\, i \in b}} m^*_{b \to i}(x_i).$$
(2.24)

and

$$\forall a \in I, \ \forall x_i \in E_i, \quad b_i^t(x_i) \propto e^{-H_i(x_i)}$$
(2.25)

Then, for any $\alpha \in \mathscr{H}$, $b_\alpha > 0$. Fixed points of the belief propagation algorithm are in correspondence with critical points of the generalized Bethe free energy [15, 18, 25]; therefore, b is a critical point of $BT_{\mathscr{A}, \mathscr{H}}$.

$\square$

3 Background on the topology of partially ordered sets

3.1 Partially ordered spaces as topological spaces

A poset $\mathscr{A}$ is a set with a binary relation, denoted as $\cdot \leq \cdot$, that is transitive ($b \leq a$ and $c \leq b$ then $c \leq a$), reflexive ($a \leq a$), and antisymmetric ($a \leq b$ and $b \leq a$ then $a = b$). A subposet of $\mathscr{A}$ is a subset $\mathscr{B}$ equipped with the same relation $\leq$. An order preserving map (or isotone map) $\phi : \mathscr{A} \to \mathscr{B}$ between two posets $\mathscr{A}$ and $\mathscr{B}$ is a function such that whenever $b \leq a$ then $\phi(b) \leq \phi(a)$. We will call a function $\phi : \mathscr{A} \to \mathscr{B}$ decreasing when $\phi(a) \leq \phi(b)$ whenever $b \leq a$. In what follows, when referring to maps between two posets we implicitly mean nondecreasing ones. A finite poset can be seen as a finite topological space [19], for the topology generated by neighborhoods $U_a = \{b : b \leq a\}$ (see Appendix B); the open subsets $U \subseteq \mathscr{A}$ are the lower-sets, i.e., sets such that whenever $a \in U$ then if $b \leq a$ then $b \in U$; we denote the associated topological space as $X_\mathscr{A}$. In fact, the reverse is true: any finite topological space that is T_0 is homeomorphic to a poset [19]. Continuous functions from $X_\mathscr{A} \to X_\mathscr{A}$ are in correspondence with order preserving maps from $\mathscr{A} \to \mathscr{B}$.

There are two standard ways to make a partially ordered set $\mathscr{A}$ into a topological space, which we will denote by $X_\mathscr{A}$ and $X^\mathscr{A}$.

$X_\mathscr{A}$ is the topological space whose points are the elements of $\mathscr{A}$, and the topology $\mathscr{O}$ on $X_\mathscr{A}$ is generated by the basis of neighborhoods $U_a = \{b \in \mathscr{A} \mid b \leq a\}$ for each $a \in \mathscr{A}$. One says that a subset $U \subseteq \mathscr{A}$ is a *lower set* whenever, for any $a \in U$ and any $b \in \mathscr{A}$ such that $b \leq a$, it follows that $b \in U$. The set of open subsets of $X_\mathscr{A}$ is exactly the collection of lower sets of $\mathscr{A}$.

3.2 Galois connection

Loosely speaking, a Galois connection is a map between two posets $\mathscr{A}$ and $\mathscr{B}$ that admits a kind of inverse. There are several ways to present Galois connections, depending on whether one chooses an order or its opposite, and whether one starts from $\mathscr{A}$ or from $\mathscr{B}$. In the following definition of Galois connection, we adopt a convention that highlights it as a specific case of an adjunction between two categories (the posets seen as categories). In Appendix A, we discuss how to relate our convention to the one in [23], and why they give rise to equivalent definitions.

Definition 2 (Galois Connection). *Let $\mathscr{A}, \mathscr{B}$ be two finite posets, and let $g : \mathscr{A} \to \mathscr{B}$ and $f : \mathscr{B} \to \mathscr{A}$ be two order preserving maps such that:*

$$\forall a \in \mathscr{A}, b \in \mathscr{B}, \quad g(a) \leq b \iff a \leq f(b). \tag{3.1}$$

Then we say that f, g form a Galois connection between $\mathscr{A}$ and $\mathscr{B}$ and denote it as $g \dashv f$.

Proposition 2 (Theorem 4.1 [23]). *Let $\mathscr{A}, \mathscr{B}$ be two posets, and let $(g : \mathscr{A} \to \mathscr{B}, f : \mathscr{B} \to \mathscr{A})$ be a Galois connection between $\mathscr{A}$ and $\mathscr{B}$, i.e., $g \dashv f$. Then,*

$$\forall a \in \mathscr{B}, b \in \mathscr{A}, \quad \sum_{b':f(b')=b} \mu_\mathscr{B}(a, b') = \sum_{a':g(a')=a} \mu_\mathscr{A}(a', b). \tag{3.2}$$

Therefore,

$$c_\mathscr{A}(a) = \sum_{a' : f(a')=a} c_\mathscr{B}(a'). \tag{3.3}$$

Proof. Equation 3.2 is a restatement of Theorem 4.1 of [23]. To prove equation 3.3, sum on the left and right hand of Equation 3.2 by $\sum_{a \in \mathscr{B}}$,

$$\sum_{a \in \mathscr{B}} \sum_{b':f(b')=b} \mu_{\mathscr{B}}(a,b') = \sum_{a \in \mathscr{B}} \sum_{a':g(a')=a} \mu_{\mathscr{A}}(a',b)$$

then,

$$\sum_{b':f(b')=b} \sum_{a \in \mathscr{B}} \mu_{\mathscr{B}}(a,b') = \sum_{a' \in \mathscr{A}} \mu_{\mathscr{A}}(a',b) = c_{\mathscr{A}}(b)$$

Therefore, $\sum_{b':f(b')=b} c_{\mathscr{B}}(b') = c_{\mathscr{A}}(b)$.

$\square$

3.3 Homotopic posets and core of a poset

As we have seen in the previous sections, posets can be viewed as topological spaces; hence one can make precise the claim that they are homotopy equivalent (as topological spaces); this is what we do in the next definition. We also recall the notion of a deformation retract, since we use it extensively in the main result of this paper.

Definition 3 (Homotopy of maps and posets). *Two order preserving maps $f, g : \mathscr{A} \to \mathscr{B}$ are said to be homotopic if there is a continuous map $h : X_{\mathscr{A}} \times [0,1] \to X_{\mathscr{B}}$ such that for any $a \in \mathscr{A}$, $h(a,0) = f(a)$ and $h(a,1) = g(a)$; we then denote $f \simeq g$; the topology on $[0,1]$ is the usual Euclidean one.*

Let $\mathscr{B}$ be a subposet of $\mathscr{A}$ and denote $i : \mathscr{B} \hookrightarrow \mathscr{A}$ as the inclusion. $\mathscr{B}$ is a deformation retract of $\mathscr{A}$ if there is a map $r : \mathscr{A} \to \mathscr{B}$ such that:

- *(retraction condition) $r \circ i = id$*

- *(homotopy condition) $i \circ r \simeq id$*

A strong deformation retract imposes that the homotopy $h : X_{\mathscr{A}} \times [0,1] \to X_{\mathscr{A}}$ leaves $\mathscr{B}$ invariant, i.e., for all $b \in \mathscr{B}$ and $t \in [0,1]$, $h(b,t) = b$.

More generally, two posets $\mathscr{A}, \mathscr{B}$ are said to be homotopic if there is a pair of order preserving maps $\phi : \mathscr{A} \to \mathscr{B}$, $\psi : \mathscr{B} \to \mathscr{A}$ such that $\psi \circ \phi \simeq id$ and $\phi \circ \psi \simeq id$.

In [19], a classification of posets up to homotopy is given; it requires introducing *linear* and *colinear* points, also known as up-beat and down-beat points (see [1] Section 1.3). The homotopy type of a poset $\mathscr{A}$—that is, whether two posets are homotopic or not—is fully characterized by its core. The core of a poset is a subposet, denoted $co\mathscr{A}$, that is homotopy equivalent to $\mathscr{A}$ and has minimal cardinality. Furthermore, the retractions of linear and colinear points are the two operations that allow one to reduce a poset to its core through a sequence of homotopy equivalences.

Definition 4 (Linear and colinear points). *Let $\mathscr{A}$ be a finite poset. Then, $a \in \mathscr{A}$ is linear when there is $a \uparrow > a$, such that*

$$\forall b \in \mathscr{A}, \quad b \geq a \implies b \geq a \uparrow. \tag{L}$$

10

Dually, a point a is colinear when there exists $a\downarrow < a$ such that

$$\forall b \in \mathscr{A}, \quad b \le a \implies b \le a\downarrow. \tag{coL}$$

Lemma 1. *Let $\mathscr{A}, \mathscr{B}$ be two finite posets, and let $f, g : \mathscr{A} \to \mathscr{B}$ be two order-preserving maps. If for all $a \in \mathscr{A}$, we have $f(a) \le g(a)$, then f is homotopic to g.*

Proof. The posets $\mathscr{A}$ and $\mathscr{B}$ can be viewed as finite topological spaces endowed with the lower Alexandrov topology (generated by lower sets), Corollary 3 of [19] therefore applies. □

Proposition 3. *Let $\mathscr{A}$ be a finite poset and let a be a linear point. Denote by*

$$r_{a\uparrow} : \mathscr{A} \to \mathscr{A} \setminus \{a\}$$

the retraction that sends any $b \ne a$ to itself and sends a to $a\uparrow$. Then $r_{a\uparrow}$ is a deformation retraction of $\mathscr{A}$ onto $\mathscr{A} \setminus \{a\}$.

Similarly, let a be a colinear point of $\mathscr{A}$, and let

$$r_{a\downarrow} : \mathscr{A} \to \mathscr{A} \setminus \{a\}$$

be the retraction that sends $b \ne a$ to itself and sends a to $a\downarrow$. Then $r_{a\downarrow}$ is a deformation retraction of $\mathscr{A}$ onto $\mathscr{A} \setminus \{a\}$.

Furthermore, when a is a linear point, $r_{a\uparrow}$ is left adjoint to the inclusion map $i : \mathscr{A} \setminus \{a\} \to \mathscr{A}$, i.e., $r_{a\uparrow} \dashv i$. When a is colinear, the inclusion map is right adjoint to the deformation retraction $r_{a\downarrow}$, i.e., $i \dashv r_{a\downarrow}$.

Proof. In both cases, when a is a linear or colinear point, we have $r_{a\uparrow} \circ i = \mathrm{id}$ and $r_{a\downarrow} \circ i = \mathrm{id}$. By definition, $i \circ r_{a\uparrow} \ge \mathrm{id}$ and $i \circ r_{a\downarrow} \le \mathrm{id}$. Therefore, by Lemma 1, $i \circ r_{a\uparrow}$ is homotopic to id, and $i \circ r_{a\downarrow}$ is also homotopic to id.

This shows that $r_{a\uparrow}$ and $r_{a\downarrow}$ are both deformation retractions onto $\mathscr{A} \setminus \{a\}$.

The fact that $r_{a\uparrow} \dashv i$ and $i \dashv r_{a\downarrow}$ is a consequence of Proposition 9 Appendix A. Let us explain the proof, following Definition 2 Appendic A, when a is a linear points, Equation L can be restated as,

$$\forall b \in \mathscr{A} \setminus \{a\}, \quad r_{a\uparrow}(a) \le b \iff a \le i(b) \tag{3.4}$$

For $b \ne a$, $r_{a\uparrow}(c) = c$ and $c \le b$ is of course equivalent to $c \le b$. Therefore $r_{a\uparrow} \dashv i$.

When a is a colinear point, Equation coL can be restated as,

$$\forall b \in \mathscr{A} \setminus \{a\}, \quad i(b) \le a \iff b \le r_{a\downarrow}(a) \tag{3.5}$$

Therefore $i \dashv r_{a\downarrow}$. □

Definition 5 (Core of a poset [19])**.** *The core of a poset $\mathscr{A}$ is a sub-poset $\mathscr{B} \subseteq \mathscr{A}$ such that $\mathscr{B}$ has no linear or colinear points (in $\mathscr{B}$) and such that $\mathscr{B}$ is a strong deformation retract of $\mathscr{A}$. We will denote the core of $\mathscr{A}$ as $co\mathscr{A}$.*

Proposition 4. *Any finite poset $\mathscr{A}$ has a core, and a strong deformation retract $r : \mathscr{A} \to co\mathscr{A}$ is obtained through a sequence of up and down retractions. Two posets are homotopic if and only if their cores are isomorphic.*

Proof. Proof of Theorem 2 in [19]. □

4 Minima and Critical Points of the Bethe Free Energy under Deformation Retractions

In this section we explain how one can, by keeping track of the Hamiltonians, reduce optimization of the Bethe free energy over a factor to optimizing it over its core (when regarded as a poset). We do it iteratively by retracting linear and colinear points.

By a chain in a poset $\mathscr{A}$, we mean a strictly increasing sequence $a_1 < a_2 < \cdots < a_n$, with $a_i \in \mathscr{A}$ for $i = 1, \ldots, n$.

Let $\mathscr{H} = (I, A)$ be a hypergraph on a finite set I. We first remark that any linear point in a factor graph is a minimal element of $\mathscr{A}(H)$: the only chains of $\mathscr{A}(H)$ are constitutes of vertices $i \in I$ and $a \in A$ such that $i < a$ (i.e. $i \in a$) or i itself. The colinear points are maximal elements. All chains have length at most 1. After removing a linear or a colinear point, the chains of the resulting subposet are still of length at most 1.

After retracting a linear point $a \in \mathscr{A}(\mathscr{H})$, it may happen that $\mathscr{A} \setminus \{a\}$ is no longer the poset associated to any factor graph. For example, consider the hypergraph $\mathscr{H} = \big(I = \{i\}, A = \{\{i\}\}\big)$. Then $i \subseteq \{i\}$, so i is a linear point. After the retraction, the poset contains only the single element $a = \{i\}$, with $E_a = E_i$. But by definition there is no factor graph having $I = \varnothing$ and E_i nonempty.

The previous remark is not a major issue: as explained in Section 2.2, the Bethe free energy is defined for subcollections $\mathscr{A}$ of subsets of a finite set I, where for each $a \in \mathscr{A}$ one defines $E_a = \prod_{i \in a} E_i$. After retraction one still associates to each $b \in \mathscr{A} \setminus \{a\}$ the finite set $E_b = \prod_{i \in b} E_i$. The poset obtained after retraction, $\mathscr{A} \setminus \{a\}$, has maximal chains of length at most 1.

In the next subsection we prove the isomorphism on critical points of the respective Bethe free energies before and after retraction, for finite posets that are collections of subsets of a finite set I and whose chains have length at most one.

4.1 Under the retraction of a linear point

Proposition 5. *Let I be a finite set, and let $\mathscr{A} \subseteq \mathscr{P}(I)$ be a poset (under inclusion) whose chains have length at most 1. Let E_i for $i \in I$ be a collection of finite sets and consider a collection of Hamiltonians $\big(H_c : E_c \to \mathbb{R}\big)_{c \in \mathscr{A}(\mathscr{H})}$. Let $a \in \mathscr{A}$ be a linear point, and set $\mathscr{B} = \mathscr{A} \setminus \{a\}$. Let $E_b = \prod_{i \in b} E_i$, let $H_b : E_b \to \mathbb{R}$ be a collection of Hamiltonians, one for each $b \in \mathscr{A}$. Define $i^* H = (H_b)_{b \in \mathscr{B}}$, where I and $(E_i)_{i \in I}$ remain unchanged.*

Then the inclusion $i : \mathscr{A} \setminus \{a\} \hookrightarrow \mathscr{A}$ induces a bijection $\mathbb{L}(\mathscr{A}) \to \mathbb{L}(\mathscr{B})$, which restricts to a bijection between the critical points of the Bethe free energy functional $BT_{\mathscr{A}, H} : \mathbb{L}(\mathscr{A})^\circ \to \mathbb{R}$ and the critical points of the Bethe free energy functional $BT_{\mathscr{B}, i^* H} : \mathbb{L}(\mathscr{B})^\circ \to \mathbb{R}$. A similar bijection holds for the global minima over $\mathbb{L}(\mathscr{A})$ and $\mathbb{L}(\mathscr{B})$.

Proof. Let us first show that $\mathbb{L}(\mathscr{A})$ is in (linear) bijection with $\mathbb{L}(\mathscr{B})$. Define two maps

$$\phi : \prod_{c \in \mathscr{A}} \mathbb{P}(E_c) \ \to \ \prod_{c \in \mathscr{B}} \mathbb{P}(E_c), \qquad s : \prod_{c \in \mathscr{B}} \mathbb{P}(E_c) \ \to \ \prod_{c \in \mathscr{A}} \mathbb{P}(E_c),$$

defined respectively as follows:

$$\forall b = (b_c)_{c \in \mathscr{A}}, \ \forall c \in \mathscr{B}, \quad \phi(b)_c = b_c,$$

$$\forall b = (b_c)_{c \in \mathscr{B}}, \ \forall c \in \mathscr{B}, \quad s(b)_c = b_c, \qquad \text{and} \qquad \forall x \in E_a, \quad s(b)_a(x) = \sum_{\substack{y \in E_{a\uparrow} \\ y_a = x}} b_{a\uparrow}(y). \quad (4.1)$$

Then one shows that : $s \circ \phi = \text{id}$ and $\phi \circ s = \text{id}$.

By Proposition 2 as $a \notin \text{im}\, i$, then $c_{\mathscr{A}}(a) = 0$; and for $b \in \mathscr{A}$, such that $b \neq a$, $c_{\mathscr{A}}(b) = c_{\mathscr{B}}(b)$. Denote $(\mathbb{E}_{Q_b}[H_b] - S(Q_b))$ as $l_b(Q_b)$.

Then, for any $Q \in \prod_{b \in \mathscr{A}} \mathbb{P}(E_b)$,

$$\sum_{b \in \mathscr{A}} c_{\mathscr{A}}(b) l_b(Q_b) = \sum_{b \in \mathscr{B}} c_{\mathscr{A}}(b) l_b(Q_b). \qquad (4.2)$$

And so, for any $Q \in \mathbb{L}(\mathscr{A})$,

$$\mathscr{L}_{\mathscr{A}}(Q) = \mathscr{L}_{\mathscr{B}}(\phi(Q)) \qquad (4.3)$$

In particular, since ϕ is a bijection, the global minima of $\mathscr{L}_{\mathscr{A}}$ are mapped bijectively to those of $\mathscr{L}_{\mathscr{B}}$. Let us now derive a similar result for critical points in $\mathbb{L}(\mathscr{A})^\circ$ and $\mathbb{L}(\mathscr{B})^\circ$.

We will now differentiate Equation 4.3 to obtain the desired property regarding the critical points of the two functionals.

Let us denote $T\mathbb{L}(\mathscr{A})$ the underlying vector space of the affine space $\mathbb{L}(\mathscr{A})$ which is the set,

$$T\mathbb{L}(\mathscr{A}) = \Big\{ u = (u_a)_{a \in \mathscr{A}} \in \prod_{a \in \mathscr{A}} \mathbb{R}^{E_a} \ \Big| \ \forall c \leq b, \ \forall x_c \in E_c, \ u_c(x_c) = \sum_{\substack{y_b \in E_b \\ y_c = x_c}} u_b(y_b),$$

$$\forall a \in \mathscr{A}, \ \sum_{x \in E_a} u_a(x) = 0 \Big\}. \qquad (4.4)$$

For any $u \in T\mathscr{L}(\mathscr{A})$, and $Q \in \mathscr{L}(\mathscr{A})^\circ$,

$$d_Q \mathscr{L}_{\mathscr{A}}(u) = d_{\phi(Q)} \mathscr{L}_{\mathscr{B}} \circ d_Q \phi(u) \qquad (4.5)$$

ϕ is an isomorphism, so its differential is invertible, and the vanishing of $d_Q \mathscr{L}_{\mathscr{A}}$ is equivalent to that of $d_{\phi(Q)} \mathscr{L}_{\mathscr{B}}$, which completes the proof.

$\square$

4.2 Under retraction of colinear points

Definition 6. *Let I be a finite set, and let $\mathscr{A} \subseteq \mathscr{P}(I)$ be a poset (under inclusion) whose chains have length at most 1. Let a be a colinear point of $\mathscr{A}$. Let $(H_c : E_c \to \mathbb{R}; c \in \mathscr{A})$ be a collection of Hamiltonians. Let $\mathscr{B} = \mathscr{A} \setminus \{a\}$ and i be the inclusion $\mathscr{B} \hookrightarrow \mathscr{A}$. Define $\psi^{a,H} : \mathbb{L}(\mathscr{B}) \to \mathbb{L}(\mathscr{A})$, denoted simply as ψ when clear from context, as the map defined as follows: for any $Q \in \mathbb{L}(\mathscr{B})$,*

$$\forall b \neq a \quad \psi(Q)_b = Q_b, \tag{4.6}$$

$$\forall x \in E_a, \quad \psi(Q)_a(x) = \pi(x \mid x_{a\downarrow}) \, Q_{a\downarrow}(x_{a\downarrow}). \tag{4.7}$$

where, for any $x_a \in E_a$,

$$\pi\left(x_a \mid x_{a\downarrow}\right) = \frac{e^{-H_a(x_a)}}{\displaystyle\sum_{\substack{z \in E_a \\ z_{a\downarrow} = x_{a\downarrow}}} e^{-H_a(z)}}. \tag{4.8}$$

Proposition 6. *Let I be a finite set, and let $\mathscr{A} \subseteq \mathscr{P}(I)$ be a poset (under inclusion) whose chains have length at most 1. Let E_i for $i \in I$ be a collection of finite sets and consider a collection of Hamiltonians $\left(H_c : E_c \to \mathbb{R}\right)_{c \in \mathscr{A}(\mathscr{H})}$. Let a be a colinear point of $\mathscr{A}$. Let $\mathscr{B} = \mathscr{A} \setminus \{a\}$ and i be the inclusion $\mathscr{B} \hookrightarrow \mathscr{A}$. Let,*

$$\forall Q \in \mathbb{L}(\mathscr{B}), \quad \mathscr{L}(Q) = \sum_{b \in \mathscr{B}} c_{\mathscr{B}}(b) \left(\mathbb{E}_{Q_b}[H_b] - S(Q_b)\right) - \mathbb{E}_{Q_{a\downarrow}}\left[\ln \sum_{\substack{z \in E_a \\ z_{a\downarrow} = X_{a\downarrow}}} e^{-H_a(z) + H_{a\downarrow}(X_{a\downarrow})}\right] \tag{4.9}$$

Then, ψ induces an isomorphism between the the critical points (respectively, the minima when they exists) of $\mathscr{L} : \mathbb{L}(\mathscr{B}) \to \mathbb{R}$ and the critical points (respectively the minima when they exist) of $BT_{\mathscr{A},H} : \mathbb{L}(\mathscr{A}) \to \mathbb{R}$.

Proof. Let $g : \mathscr{A} \to \mathscr{B}$ and $f : \mathscr{B} \to \mathscr{A}$ such that g, f for a Galois connection, i.e. $f \dashv g$. By Proposition 2,

$$\forall Q \in \prod_{a \in \mathscr{A}} \mathbb{P}(E_a), \quad \sum_{a \in \mathscr{A}} c_{\mathscr{A}}(a) \left[\mathbb{E}_{Q_a}[H_a] - S(Q_a)\right] = \sum_{a' \in \mathscr{B}} \sum_{a \in \mathscr{A} : g(a) = a'} c_{\mathscr{A}}(a) \left[\mathbb{E}_{Q_a}[H_a] - S(Q_a)\right]. \tag{4.10}$$

Let $a \in \mathscr{A}$ be a colinear point, let $\mathscr{B} = \mathscr{A} \setminus \{a\}$. Then $c_{\mathscr{A}}(b) = c_{\mathscr{B}}(b)$ for any $b \in \mathscr{B}$ such that $b \neq a \downarrow$, and,

$$c_{\mathscr{B}}(a \downarrow) = c_{\mathscr{A}}(a \downarrow) + c_{\mathscr{A}}(a) \tag{4.11}$$

Let $g = r_{a\downarrow} : \mathscr{A} \to \mathscr{B}$ be the retraction of a on $a \downarrow$. The next step is to apply the chain rule for entropy to the joint random variables (X, Y), where for $x \in E_a$ we set $X(x) = x$ and $Y(x) = x_{a\downarrow}$. Let $Q_a \in \mathbb{P}(F_a)$, let us denote $Q_{a\downarrow} := \pi^a_{a\downarrow,*} Q_a$ (the marginalisation induced by the projection onto $E_{a\downarrow}$: $\pi^a_{a\downarrow}$) and let,

$$\forall x_a \in E_a, y_{a\downarrow} \in E_{a\downarrow}, \quad Q_{a|a\downarrow}(x_a, y_{a\downarrow}) = \frac{Q_a(x_a)}{Q_{a\downarrow}(y_{a\downarrow})} 1[x_{a\downarrow} = y_{a\downarrow}].$$

Then the following holds,

$$\forall Q_a \in \mathbb{P}(E_a), \quad S(Q_a) = S(Q_{a\downarrow}) + \mathbb{E}_{Q_{a\downarrow}}[S(Q_{a|a\downarrow})]. \tag{4.12}$$

14

where, more explicitly,

$$S(Q_{a|a\downarrow})(y_{a\downarrow}) = - \sum_{x_a \, : \, x_{a\downarrow} = y_{a\downarrow}} Q_{a|a\downarrow}(x_a, y_{a\downarrow}) \ln Q_{a|a\downarrow}(x_a, y_{a\downarrow}). \tag{4.13}$$

Therefore, for any $Q_a \in \mathbb{P}(E_a)$,

$$c_{\mathscr{A}}(a) \left(\mathbb{E}_{Q_a}[H_a] - S(Q_a) \right) + c_{\mathscr{A}}(a \downarrow) \left(\mathbb{E}_{Q_{a\downarrow}}[H_{a\downarrow}] - S(Q_{a\downarrow}) \right)$$
$$= (c_{\mathscr{A}}(a) + c_{\mathscr{A}}(a \downarrow)) \left(\mathbb{E}_{Q_{a\downarrow}}[H_{a\downarrow}] - S(Q_{a\downarrow}) \right) + c_{\mathscr{A}}(a) \left(\mathbb{E}_{Q_a}[H_a - H_{a\downarrow} \circ \pi^a_{a\downarrow}] - \mathbb{E}_{Q_{a\downarrow}}[S(Q_{a|a\downarrow})] \right). \tag{4.14}$$

First part: proof for global minima.

Let us define, for any $P \in \mathbb{L}(\mathscr{B})$,

$$\mathscr{L}(P) = \min_{Q_a \in \mathbb{P}(E_a) \, : \, \pi^a_{a\downarrow,*}(Q_a) = P_{a\downarrow}} BT_{\mathscr{A},H}(P, Q_a). \tag{4.15}$$

Then

$$\min_{P \in \mathbb{L}(\mathscr{A})} BT_{\mathscr{A},H}(p) = \min_{P \in \mathbb{L}(\mathscr{B})} \mathscr{L}(P) \tag{4.16}$$

One remarks that,

$$\forall P \in \mathbb{L}(\mathscr{B}), \quad \mathscr{L}(P) = \sum_{b \in \mathscr{B}} c_{\mathscr{B}}(b) \left(\mathbb{E}_{P_b}[H_b] - S(P_b) \right)$$
$$+ c_{\mathscr{A}}(a) \left(\min_{\substack{Q_a \in \mathbb{P}(E_a) \\ \pi^a_{a\downarrow,*}(Q_a) = P_{a\downarrow}}} \mathbb{E}_{Q_a}[H_a - H_{a\downarrow} \circ \pi^a_{a\downarrow}] - \mathbb{E}_{Q_{a\downarrow}}[S(Q_{a|a\downarrow})] \right). \tag{4.17}$$

In particular as a is maximal in $\mathscr{A}$ then $c(a) = 1$. The sub-optimization problem to solve is the following:

$$\min_{Q_a \in \mathbb{P}(E_a) \, : \, \pi^a_{a\downarrow,*}(Q_a) = P_{a\downarrow}} \mathbb{E}_{Q_a}[H_a - H_{a\downarrow} \circ \pi^a_{a\downarrow}] - \mathbb{E}_{P_{a\downarrow}}[S(q_{a|a\downarrow})] \tag{4.18}$$

The parametrization of the constraint is given by, for any $x_a \in E_a$, with $x_{a\downarrow} = \pi^a_{a\downarrow}(x_a)$,

$$Q_a(x_a) = \pi(x_a \mid x_{a\downarrow}) P_{a\downarrow}(x_{a\downarrow}).$$

In particular, for any $x \in E_a$ and $y \in E_{a\downarrow}$, $\pi(x \mid y) = Q_{a|a\downarrow}(x_a, x_{a\downarrow}) 1[x_{a\downarrow} = y]$. Pose $f_a(x_a) = H_a(x_a) - H_{a\downarrow}(\pi^a_{a\downarrow}(x_a))$, in other words $f_a = H_a - H_{a\downarrow} \circ \pi^a_{a\downarrow}$, then,

$$\mathbb{E}_{Q_a}[f_a] - \mathbb{E}_{P_{a\downarrow}}[S(Q_{a|a\downarrow})] = \sum_{y \in E_{a\downarrow}} P_{a\downarrow}(y) \sum_{x_a \, : \, \pi^a_{a\downarrow}(x_a) = y} \pi(x_a|y) \ln \frac{\pi(x_a|y)}{e^{-f_a(x_a)}} \tag{4.19}$$

In Equation 4.19, for each $y \in E_{a\downarrow}$, the map

$$\pi(\cdot|y) \to \sum_{x_a : \pi^a_{a\downarrow}(x_a)=y} \pi(x_a, y) \ln \frac{\pi(x_a|y)}{e^{-f_a}(x_a)}$$

is minimal for,

$$\forall x \in E_a, \quad \pi^*(x \mid y) = \frac{e^{-f_a(x)}}{\displaystyle\sum_{z \in E_a : z_{a\downarrow}=y} e^{-f_a(z)}} \cdot 1[x_{a\downarrow} = y]. \tag{4.20}$$

Denote by $\mathbb{P}_{>0}(E_a \mid E_{a\downarrow})$ the set of kernels $y \in E_{a\downarrow} \to \pi(\cdot \mid y) \in \mathbb{P}(E_a)$ such that:

1) if $x_{a\downarrow} \neq y$, then $\pi(x \mid y) = 0$.

2) for all $x \in E_a$, $\pi(x \mid x_{a\downarrow}) > 0$.

$C : \pi \in \mathbb{P}_{>0}(E_a \mid E_{a\downarrow}) \to \mathbb{E}_{Q_a}[f_a] - \mathbb{E}_{P_{a\downarrow}}[S(Q_{a|a\downarrow})]$ is a strictly convex function bounded by below on $\mathbb{P}_{>0}(E_a \mid E_{a\downarrow})$, therefore it's minimum is unique and attained. Therefore the global minimum of C is attained for $\pi^* : E_{a\downarrow} \to \mathbb{P}(E_a)$ defined as, for any $x \in E_a$ and $y \in E_{a\downarrow}$,

$$\pi^*(x \mid y) = \frac{e^{-H_a(x)}}{\displaystyle\sum_{z \in E_a : z_{a\downarrow}=y} e^{-H_a(z_a)}} \cdot 1[x_{a\downarrow} = y].$$

The associated optimal Q_a^* is such that for any $x \in E_a$, $Q_a^*(x) = \pi^*(x \mid x_{a\downarrow}) \cdot P_{a\downarrow}(x_{a\downarrow})$. Therefore $(P, Q_a^*) = \phi(Q)$.

In particular,

$$\max_{Q_a \in \mathbb{P}(E_a) : \pi^a_{a\downarrow,*}(Q_a)=P_{a\downarrow}} \mathbb{E}_{Q_a}[H_a - H_a \circ \pi^a_{a\downarrow}] + \mathbb{E}_{P_{a\downarrow}}[S(Q_{a|a\downarrow})] = -\mathbb{E}_{P_{a\downarrow}}\left[\ln \sum_{z \in E_a : z_{a\downarrow}=X_{a\downarrow}} e^{-H_a(z)+H_{a\downarrow}(X_{a\downarrow})} \right]$$

$$= - \sum_{y \in E_{a\downarrow}} P_{a\downarrow}(y) \ln \sum_{z \in E_a : z_{a\downarrow}=y} e^{-H_a(z)+H_{a\downarrow}(y)} \tag{4.21}$$

We have solved the sub-optimization problem and therefore obtained an analytic expression for $\mathscr{L}$,

$$\forall Q \in \mathbb{L}(\mathscr{B}), \quad \mathscr{L}(Q) = \sum_{b \in \mathscr{B}} c_{\mathscr{B}}(b) \left(\mathbb{E}_{Q_b}[H_b] - S(Q_b) \right) - \mathbb{E}_{Q_{a\downarrow}}\left[\ln \sum_{z : z_{a\downarrow}=X_{a\downarrow}} e^{-H_a(z)+H_{a\downarrow}(X_{a\downarrow})} \right] \tag{4.22}$$

Therefore, ϕ induces an isomorphism between the minima of $\mathscr{L}$ and $BT_{\mathscr{A},H}$.

<u>Second part: proof for critical points.</u>

Recall that

$$\forall Q \in \mathbb{L}(\mathscr{A}), \quad BP_{\mathscr{A},H}(Q) = \sum_{b \in \mathscr{B}} c_{\mathscr{B}}(b) \left(\mathbb{E}_{Q_b}[H_b] - S(Q_b) \right)$$
$$+ c_{\mathscr{A}}(a) \left(\mathbb{E}_{Q_a}[H_a - H_{a\downarrow} \circ \pi^a_{a\downarrow}] - \mathbb{E}_{Q_{a\downarrow}}[S(Q_{a|a\downarrow})] \right). \tag{4.23}$$

The constraint $Q \in \mathbb{L}(\mathscr{A})^\circ$ can be rewritten as $(Q_b)_{b \in \mathscr{B}} \in \mathbb{L}(\mathscr{B})^\circ$, and $\pi^a_{a\downarrow} Q_a = Q_{a\downarrow}$. The second constraint can be reparameterized as

$$\forall x \in E_a, \quad Q_a(x) = \pi(x \mid x_{a\downarrow}) Q_{a\downarrow}(x_{a\downarrow}).$$

with $\pi \in \mathbb{P}_{>0}(E_a \mid E_{a\downarrow})$. Therefore, $\mathbb{L}(\mathscr{A})^\circ \cong \mathbb{L}(\mathscr{B})^\circ \times \mathbb{P}_{>0}(E_a \mid E_{a\downarrow})$. Both $\mathbb{L}(\mathscr{B})^\circ$ and $\mathbb{P}_{>0}(E_a \mid E_{a\downarrow})$ are open subsets of affine spaces, and the underlying vector space of their product, i.e. $T\mathbb{L}(\mathscr{A})$, is the product $T\mathbb{L}(\mathscr{B}) \times T\mathbb{P}_{>0}(E_a \mid E_{a\downarrow})$. For any $(v, w) \in T\mathbb{L}(\mathscr{A})$, and $Q = (P, Q_a) \in \mathbb{L}(\mathscr{A})^\circ$,

$$d_Q BT_{\mathscr{A},H}(v, w) = d_P BT_{\mathscr{B},H}(v)$$
$$+ \sum_{y \in E_{a\downarrow}} v_{a\downarrow}(y) \sum_{\substack{x \in E_a \\ x_{a\downarrow}=y}} \pi(x \mid y) \ln \frac{\pi(x \mid y)}{e^{-f_a(x)}}$$
$$+ \sum_{y \in E_{a\downarrow}} P_{a\downarrow}(y) \sum_{\substack{x \in E_a \\ x_{a\downarrow}=y}} w(x \mid y) \left(\ln \frac{\pi(x \mid y)}{e^{-f_a(x)}} + 1 \right) \tag{4.24}$$

Therefore $d_Q BT_{\mathscr{A},H}(v, w) = 0$ for all $(v, w) \in T\mathscr{L}(\mathscr{B}) \times T\mathbb{P}_{>0}(E_a | E_{a\downarrow})$ is equivalent to, for any v, w,

$$d_P BT_{\mathscr{B},H}(v) + \sum_{y \in E_{a\downarrow}} v_{a\downarrow}(y) \sum_{x : x_{a\downarrow}=y} \pi(x|y) \ln \frac{\pi(x|y)}{e^{-f_a(x)}} = 0 \tag{4.25}$$

$$\sum_{x \in E_a : x_{a\downarrow}=y} w(x|y) \left(\ln \frac{\pi(x|y)}{e^{-f_a(x)}} + 1 \right) = 0 \tag{4.26}$$

The solution to Equation 4.26 is given by,

$$\forall y \in E_{a\downarrow}, \forall x \in E_a, \quad \pi^*(x|y) = \frac{e^{-f_a(x)}}{\sum_{x : x_{a\downarrow}=y} e^{-f_a(x)}} \cdot \mathbb{1}[x_{a\downarrow} = y], \tag{4.27}$$

and Equation 4.25 can be rewritten as, for any $v \in T \lim \mathbb{L}(\mathscr{B})$,

$$d_P \mathscr{L}(v) = 0 \tag{4.28}$$

Then ψ is an isomorphism between the critical points of $BT_{\mathscr{A},H}$ and $\mathscr{L}$.

$\square$

In the last part of this section, we rewrite $\mathscr{L}$ as $BT_{\mathscr{B},\widetilde{H}}$ with an appropriate choice of Hamiltonians $\widetilde{H}$.

Proposition 7. *Let I be a finite set, and let $\mathscr{A} \subseteq \mathscr{P}(I)$ be a poset (under inclusion) whose chains have length at most 1. Let E_i for $i \in I$ be a collection of finite sets and consider a collection of Hamiltonians $\left(H_c : E_c \to \mathbb{R}\right)_{c \in \mathscr{A}(\mathscr{H})}$. Let a be a colinear point of $\mathscr{A}$. Let $\mathscr{B} = \mathscr{A} \setminus \{a\}$, and let $i : \mathscr{B} \hookrightarrow \mathscr{A}$ be the inclusion map. There are two possible cases:*

1. *if $b = a \downarrow$ is a linear point of $\mathscr{B}$, then pose for any $c \in \mathscr{B}$ such that $c \neq b \uparrow$, $\widetilde{H}_b = H_b$ and*

$$\forall x \in E_b, \quad \widehat{H}_b(x) = \ln \sum_{z \in E_a \,:\, z_b = x} e^{-H_a(z) + H_b(x)}$$

$$\widetilde{H}_{b\uparrow} = H_{b\uparrow} - \widehat{H}_b \circ \pi_b^{b\uparrow} \tag{4.29}$$

2. *if $a \downarrow$ is not a linear point of $\mathscr{B}$, then, for any $b \in \mathscr{B}$ such that $b \neq a \downarrow$, $\widetilde{H}_b = H_b$, and*

$$\forall x \in E_a, \quad \widetilde{H}_{a\downarrow}(x) = H_{a\downarrow}(x) \; - \; \frac{1}{c_{\mathscr{B}}(a\downarrow)} \left[\ln \sum_{\substack{z \in E_a \\ z_{a\downarrow} = x}} e^{-H_a(z) + H_{a\downarrow}(x)} \right]. \tag{4.30}$$

Then, ψ induces an isomorphism between the minima of $BT_{\mathscr{B}, \widetilde{H}} : \mathbb{L}(\mathscr{B}) \to \mathbb{R}$ and the minima the of $BT_{\mathscr{A}, H} : \mathbb{L}(\mathscr{A}) \to \mathbb{R}$; and also the critical points of $BT_{\mathscr{B}, \widetilde{H}} : \mathbb{L}(\mathscr{B})^\circ \to \mathbb{R}$ and $BT_{\mathscr{A}, H} : \mathbb{L}(\mathscr{A})^\circ \to \mathbb{R}$.

Proof. Any element $a \in \mathscr{A}$ is either be a minimal element or a maximal element (or both). In particular, if a is a linear point, then $a \uparrow > a$, so a is a minimal element; conversely, if a is a colinear point, then $a \downarrow < a$, so a is a maximal element. Let $a \in \mathscr{A}$ be a minimal element but not a linear point, then, there are at least two elements b, b_1 such that $a < b$ and $a < b_1$; one shows that, $\mu_{\mathscr{A}}(a, a) = 1$ and for any $a < c$ (c is a cover of a, i.e. there is not $a < x < c$) $\mu(c, a) = -1$. Therefore $c_{\mathscr{A}}(a) < -1$. Therefore, when $a \downarrow$ is not a linear point of $\mathscr{B}$, by Proposition 6, the critical points and the minima of $BT_{\mathscr{A}, H}$ are in correspondence with those of $\mathscr{L}$, where

$$\forall P \in \mathbb{L}(\mathscr{B}), \quad \mathscr{L}(P) = \sum_{b \in \mathscr{B}} c_{\mathscr{B}}(b) \left(\mathbb{E}_{P_b}[H_b] - S(P_b) \right) - \mathbb{E}_{P_{a\downarrow}} \left[\ln \sum_{z \in E_a \,:\, z_{a\downarrow} = X_{a\downarrow}} e^{-H_a(z) + H_{a\downarrow}(X_{a\downarrow})} \right] \tag{4.31}$$

This expression can be rewritten as

$$\forall Q \in \mathbb{L}(\mathscr{B}), \quad \mathscr{L}(Q) = \sum_{b \in \mathscr{B}} c_{\mathscr{B}}(b) \left(\mathbb{E}_{Q_b}[\widetilde{H}_b] - S(Q_b) \right) = BT_{\mathscr{B}, \widetilde{H}}(Q). \tag{4.32}$$

Assume now that $b = a \downarrow$ is a linear point of $\mathscr{B}$, then by Proposition 2, $c_{\mathscr{B}}(b) = 0$. $b \uparrow$ is a maximal element of $\mathscr{B}$, therefore $c_{\mathscr{B}}(b \uparrow) = 1$. Therefore,

$$\forall Q \in \mathbb{L}(\mathscr{B}), \quad \mathscr{L}(Q) = \sum_{d \in \mathscr{B}: d \neq b\uparrow} c_{\mathscr{B}}(d) \left(\mathbb{E}_{Q_d}[H_d] - S(Q_d) \right)$$

$$+ \mathbb{E}_{Q_{b\uparrow}}[H_{b\uparrow}] - \mathbb{E}_{Q_{a\downarrow}} \left[\ln \sum_{z \in E_a : z_{a\downarrow} = X_{a\downarrow}} e^{-H_a(z) + H_{a\downarrow}(X_{a\downarrow})} \right] - S(Q_a) \tag{4.33}$$

Pose for all $x \in E_{a\downarrow}$,

$$\widehat{H}_{a\downarrow}(x) = \ln \sum_{z \in E_a : z_{a\downarrow} = x} e^{-H_a(z) + H_{a\downarrow}(x)} \tag{4.34}$$

One remarks that,

$$\mathbb{E}_{Q_{b\uparrow}}[H_{b\uparrow}] - \mathbb{E}_{Q_{a\downarrow}}[\widehat{H}_{a\downarrow}] = \mathbb{E}_{Q_{b\uparrow}} \left[H_{b\uparrow}(X_{b\uparrow}) - \widehat{H}_{a\downarrow} \circ \pi_{a\downarrow}^{b\uparrow}(X_{b\uparrow}) \right]$$

$$= \sum_{x \in E_{b\uparrow}} Q_{b\uparrow}(x) \left[H_{b\uparrow}(x) - \widehat{H}_{a\downarrow}(x_{a\downarrow}) \right] \tag{4.35}$$

Therefore by posing, for $c \in \mathscr{B}$ such that $c \neq b\uparrow$, $\widetilde{H}_c = H_c$ and $\widetilde{H}_{b\uparrow} = H_{b\uparrow} - \widehat{H}_{a\downarrow} \circ \pi_{a\downarrow}^{b\uparrow}$, then $\mathscr{L} = BT_{\mathscr{B}, \widetilde{H}}$.

$\square$

5 The main result: minima and critical points under deformation retractions of hypergraphs

Theorem 1. *Let $\mathscr{F}$ be a factor graph, with sets E_i for $i \in I$, hypergraph $\mathscr{H} = (I, A)$ and Hamiltonians $\left(H_c : E_c \to \mathbb{R} \right)_{c \in \mathscr{A}(\mathscr{H})}$ (associated to the factors). Denote $\mathscr{A}(\mathscr{H})$ as $\mathscr{A}$. Let $\mathrm{co}\mathscr{A}$ be the core of $\mathscr{A}$. Then, there is a a collection of functions $(\widetilde{H}_c : E_c \to \mathbb{R})_{c \in \mathrm{co}\mathscr{A}}$ that induces an isomorphism between the critical points of $BT_{\mathscr{A}, H}$ and those of $BT_{\mathrm{co}\mathscr{A}, \widetilde{H}}$.*

Proof. By Proposition 4, there exists a sequence of up and down retractions, corresponding respectively to the deletion of linear and colinear points of $\mathscr{A}$, which, when composed, define a deformation retraction from $\mathscr{A}$ onto $\mathrm{co}\mathscr{A}$. Let us denote $a_1, ... a_n$ the set of points of $\mathscr{A}$ that are successively deleted to reach $\mathrm{co}\mathscr{A}$. Each a_k, for $k = 1...n$, are linear or colinear points in their respective posets.

If a_1 is a linear point, one can apply Proposition 5. Denote $\mathscr{A}_1 = \mathscr{A} \setminus \{a_1\}$, and let $i : \mathscr{A}_1 \hookrightarrow \mathscr{A}$ be the inclusion of $\mathscr{A}_1$ into $\mathscr{A}$; let $\phi : \mathbb{L}(\mathscr{A}) \to \mathbb{L}(\mathscr{B})$ be the induced map. Denote $H_1 = i^*H$; then ϕ induces an isomorphism between the critical points of $BT_{\mathscr{A}, H}$ and those of $BT_{\mathscr{A}_1, H_1}$. Denote by g_{a_1} the restriction of ϕ to the respective sets of critical points.

If a_1 is a colinear point, then one instead applies Proposition 7. In this case, let $H_1 = \widetilde{H}$. Recall that $\psi^{a_1, H}$ (Definition 6) induces an isomorphism between the critical points of $BT_{\mathscr{A}, H}$ and those of $BT_{\mathscr{A}_1, H_1}$; denote by g_{a_1} the inverse of $\psi^{a_1, H}$, restricted to the respective critical points. The chains of $\mathscr{A}$ are of length at most 1, so are those of $\mathscr{A}_1$.

Let us build by recursion the sequences $\mathscr{A}_k$, H_k, and g_{a_k} for $1 < k \leq n$. Repeating the same argument as for a_1, and depending on whether a_k is a linear or colinear point, we set $\mathscr{A}_{k+1} = \mathscr{A}_k \setminus \{a_k\}$; the chains in $\mathscr{A}_{k+1}$ are of length at most 1 as are those of $\mathscr{A}_k$. H_{k+1} is respectively $i^* H_k$ or $\widetilde{H}_k$; and $g_{a_{k+1}}$ is respectively the restriction of $\phi : \mathbb{L}(\mathscr{A}_k) \to \mathbb{L}(\mathscr{A}_{k+1})$ to the critical points of $BT_{\mathscr{A}_k, H_k}$ and $BT_{\mathscr{A}_{k+1}, H_{k+1}}$, or the restriction of ψ^{a_k, H_k} to those critical points.

Then $g_{a_n} \circ ... \circ g_{a_1}$ defines a isomorphism between the critical points of $BT_{\mathscr{A}, H} : \mathbb{L}(\mathscr{A}) \to \mathbb{R}$ and those of $BT_{\mathrm{co}\mathscr{A}, H_{n+1}} : \mathbb{L}(\mathrm{co}\mathscr{A}) \to \mathbb{R}$.

$\square$

Acknowledgements

References

[1] Jonathan A Barmak. *Algebraic topology of finite topological spaces and applications*, volume 2032. Springer Science & Business Media, 2011.

[2] Karl-Magnus Dahlén and Christopher Lindberg. Message-passing methods on loopy factor graphs based on homotopy and reweighting, 2012.

[3] Cyril Furtlehner and Jean-Marc Lasgouttes. Loopy belief propagation inference with a prescribed fixed point.

[4] Anna Grim and Pedro Felzenszwalb. *Message Passing Dynamics of Belief Propagation Algorithms*. PhD thesis, Doctoral dissertation, Brown University, 2022.

[5] Tom Heskes. Stable fixed points of loopy belief propagation are local minima of the bethe free energy. *Advances in neural information processing systems*, 15, 2002.

[6] Alexander T Ihler, John W Fisher III, Alan S Willsky, and David Maxwell Chickering. Loopy belief propagation: convergence and effects of message errors. *Journal of Machine Learning Research*, 6(5), 2005.

[7] Christian Knoll, Dhagash Mehta, Tianran Chen, and Franz Pernkopf. Fixed points of belief propagation—an analysis via polynomial homotopy continuation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(9):2124–2136, 2017.

[8] Christian Knoll, Adrian Weller, and Franz Pernkopf. Self-guided belief propagation – a homotopy continuation method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):5139–5157, 2023.

[9] Frank R Kschischang, Brendan J Frey, and H-A Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on information theory*, 47(2):498–519, 2002.

[10] Steffen L. Lauritzen. *Graphical Models*. Oxford Science Publications, 1996.

[11] Victorin Martin, Jean-Marc Lasgouttes, and Cyril Furtlehner. Local stability of belief propagation algorithm with multiple fixed points. In *STAIRS 2012*, pages 180–191. IOS Press, 2012.

[12] Joris Mooij and Hilbert Kappen. Sufficient conditions for convergence of loopy belief propagation. *arXiv preprint arXiv:1207.1405*, 2012.

[13] Joris M. Mooij and Hilbert J. Kappen. Sufficient conditions for convergence of the sum–product algorithm. *IEEE Transactions on Information Theory*, 53(12):4422–4437, 2007.

[14] Olivier Peltre. Homology of Message-Passing Algorithms. `http://opeltre.github.io`, 2020. Ph.D. thesis (preprint).

[15] Olivier Peltre. Message passing algorithms and homology, 2020. Ph.D. thesis, Link.

[16] Olivier Peltre. Belief propagation as diffusion. In Frank Nielsen and Frédéric Barbaresco, editors, *Geometric Science of Information*, pages 547–556, Cham, 2021. Springer International Publishing.

[17] Gian-Carlo Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Probability theory and related fields*, 2(4):340–368, 1964.

[18] Grégoire Sergeant-Perthuis, Toby St Clere Smithe, and Léo Boitel. On the functoriality of belief propagation algorithms on finite partially ordered sets, 2025.

[19] Robert E Stong. Finite topological spaces. *Transactions of the American Mathematical Society*, 123(2):325–340, 1966.

[20] Sekhar C. Tatikonda and Michael I. Jordan. Loopy belief propagation and gibbs measures. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, UAI'02, page 493–500, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.

[21] Martin Wainwright, Tommi Jaakkola, and Alan Willsky. Tree consistency and bounds on the performance of the max-product algorithm and its generalizations. *Statistics and computing*, 14(2):143–166, 2004.

[22] Martin J Wainwright, Michael I Jordan, et al. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.

[23] James W. Walker. Homotopy type and euler characteristic of partially ordered sets. *European Journal of Combinatorics*, 2(4):373–384, 1981.

[24] Max Welling. On the choice of regions for generalized belief propagation. *arXiv preprint arXiv:1207.4158*, 2012.

[25] J.S. Yedidia, W.T. Freeman, and Y. Weiss. Constructing Free-Energy Approximations and Generalized Belief Propagation Algorithms. *IEEE Transactions on Information Theory*, 51(7):2282–2312, July 2005.

[26] Roie Zivan, Omer Lev, and Rotem Galiki. Beyond trees: Analysis and convergence of belief propagation in graphs with multiple cycles. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):7333–7340, Apr. 2020.

A Galois Connection and Adjunction

Definition 7 (A poset as a category)**.** *Consider a finite poset $\mathscr{A}$; the associated category, denoted as $\mathbf{A}$ or simply as $\mathscr{A}$, has as objects the elements $x \in \mathscr{A}$ and a unique morphism $x \to y$ whenever $x \leq y$.*

Definition 8 (Galois Connection)**.** *Let $\mathscr{A}, \mathscr{B}$ be two finite posets, and let $g : \mathscr{A} \to \mathscr{B}$ and $f : \mathscr{B} \to \mathscr{A}$ be two mean nondecreasing maps such that:*

$$\forall a \in \mathscr{A}, b \in \mathscr{B}, \quad g(a) \leq b \iff a \leq f(b) \tag{A.1}$$

then we say that f, g form a Galois connection between $\mathscr{A}$ and $\mathscr{B}$.

Remark 1. Definition 2 is equivalent to saying that g is left adjoint to f for $\mathscr{A}$ and $\mathscr{B}$ seen as categories. It is the reason why we denoted it in the body of the article $g \dashv f$ when f, g define a Galois connection. The relation to the Galois connection as defined in [23] is that f, g form a Galois connection in our sense if and only if f, g form a Galois connection between $\mathscr{A}^{op}$ and $\mathscr{B}$ according to the conventions of [23].

Proposition 8. *Let $\mathscr{A}, \mathscr{B}$ be two finite posets, and let $g : \mathscr{A} \to \mathscr{B}$ and $f : \mathscr{B} \to \mathscr{A}$ form a Galois connection between $\mathscr{A}$ and $\mathscr{B}$, i.e., $g \dashv f$. Then $\mathscr{A}$ and $\mathscr{B}$ have the same homotopy type.*

Proof. Section 4 [23] □

Proposition 9. *A point a is linear if and only if the inclusion map $j : \mathscr{A} \setminus \{a\}$ admits a left adjoint, denoted as r. In other words, there exists a unique map $r : \mathscr{A} \setminus \{a\} \to \mathscr{A}$ such that (r, j) is a Galois connection between $\mathscr{A}$ and $\mathscr{A} \setminus \{a\}$.*

Similarly, a point a is colinear if and only if the inclusion map $j : \mathscr{A} \setminus \{a\}$ admits a right adjoint, i.e., if (j, r) is a Galois connection between $\mathscr{A}$ and $\mathscr{A} \setminus \{a\}$.

Proof. If a is a linear point of $\mathscr{A}$, then for any $b \neq a$, we have $j \circ r(b) = b$ and $j \circ r(a) = a \uparrow \geq a \geq a$; furthermore, $r \circ j = \mathrm{id}$.

Therefore, if for $b \in \mathscr{A}$ and $c \in \mathscr{A} \setminus \{a\}$ such that $r(b) \leq c$, then

$$b \leq j \circ r(b) \leq j(c),$$

and so $b \leq j(c)$.

Now assume that $b \leq j(c)$, then $r(b) \leq r \circ j(c) = c$; therefore, $r \dashv j$. Similarly, assume a is a colinear point of $\mathscr{A}$, then it is a linear point of $\mathscr{A}^{op}$. Applying the previous result, it follows that

$$r(a) \leq_{op} b \iff a \leq_{op} f(b),$$

which corresponds to

$$b \leq_{op} r(a) \iff f(b) \leq_{op} a,$$

and so $j \dashv r$.

Now, assume that $r \dashv j$, then $a \uparrow = r(a) \leq b$ if and only if $a \leq b$, and so $a \leq b \implies a \uparrow \leq b$; therefore, a is a linear point. A similar proof shows that $j \dashv r$ implies that a is a colinear point. □

Remark 2. Let a be a linear point. Then, the left adjoint to the inclusion $j : \mathscr{A} \setminus \{a\} \to \mathscr{A}$, denoted as $r^{a\uparrow} : \mathscr{A} \to \mathscr{A} \setminus \{a\}$, sends a to $a \uparrow$ and is the identity map on the other points.

Similarly, let a be a colinear point. Then, the right adjoint to the inclusion $j : \mathscr{A} \setminus \{a\} \to \mathscr{A}$, denoted as $r^{a\downarrow} : \mathscr{A} \to \mathscr{A} \setminus \{a\}$, sends a to $a \downarrow$ and is the identity map on the other points.

B   Partially ordered spaces as topological spaces

There two standard way to make a partially ordered set $\mathscr{A}$ into a topological spaces that we will denote as $X_{\mathscr{A}}$ and $X^{\mathscr{A}}$.

$X_{\mathscr{A}}$ is the topological space whose points are the elements of $\mathscr{A}$, and the topology $\mathscr{O}$ on $X_{\mathscr{A}}$ is generated by the basis of neighborhoods $U_a = \{b \in \mathscr{A} \mid b \leq a\}$ for each $a \in \mathscr{A}$. One says that a subset $U \subseteq \mathscr{A}$ is a *lower set* whenever, for any $a \in U$ and any $b \in \mathscr{A}$ such that $b \leq a$, it follows that $b \in U$. The set of open subsets of $X_{\mathscr{A}}$ is exactly the collection of lower sets of $\mathscr{A}$.

$X^{\mathscr{A}}$ is the space $\mathscr{A}$ equipped with the topology $\mathscr{O}_1$, generated by the neighborhood bases $\{b \in \mathscr{A} \mid a \leq b\}$ for each $a \in \mathscr{A}$. The set of open subsets of $\mathscr{O}_1$ is the set of upper sets, that is, the subsets U such that if $a \in U$ and $b \in \mathscr{A}$ with $a \leq b$, then $b \in U$.

The two spaces $X_{\mathscr{A}}$ and $X^{\mathscr{A}}$ are related as follows: $X_{\mathscr{A}} = X^{\mathscr{A}^{\mathrm{op}}}$.