Submitted to:
ACT 2026

On the Functoriality of Belief Propagation Algorithms on Finite Partially Ordered Sets

Grégoire Sergeant-Perthuis[1] Toby St Clere Smithe

CQSB, Sobonne Université Kodamai Ltd

Undirected graphical models are a widely used class of probabilistic models in machine learning that capture prior knowledge or putative pairwise interactions between variables. Those interactions are encoded in a graph for pairwise interactions; however, generalizations such as factor graphs account for higher-degree interactions using hypergraphs. Inference on such models, which is performed by conditioning on some observed variables, is typically done approximately by optimizing a free energy, which is an instance of variational inference. The Belief Propagation algorithm is a dynamic programming algorithm that finds critical points of that free energy. Recent efforts have been made to unify and extend inference on graphical models and factor graphs to more expressive probabilistic models. A synthesis of these works shows that inference on graphical models, factor graphs, and their generalizations relies on the introduction of presheaves and associated invariants (homology and cohomology groups). We propose to study the impact of the transformation of the presheaves onto the associated message passing algorithms. We show that natural transformations between presheaves associated with graphical models and their generalizations, which can be understood as coherent binning of the set of values of the variables, induce morphisms between associated message-passing algorithms. It is, to our knowledge, the first result on functoriality of the Loopy Belief Propagation.

1 Introduction and related work

The Belief Propagation algorithm (BP) [1, 2, 3, 4], and its generalization, the Generalized Belief Propagation algorithm [5, 6], are celebrated for efficient approximate variational inference on graphical models, factor graphs, and more generally on models defined by factorizations over a collection of subsets of variables. They can be interpreted as inference over a diagram of marginally compatible probability distributions [7], generalizing inference from a single distribution to a collection of distributions defined on local variables that satisfy compatibility conditions.

Compositionality of probabilistic models [8], inference and variational inference [9, 10, 11], and in particular inference on acyclic graphical models such as Bayesian networks [12], have been at the center of interest in recent years.

The appearance of sheaves over combinatorial objects (cellular complexes, posets) has found applications in topological data analysis [13] and in deep learning [14, 15, 16, 17, 18]. The Belief Propagation algorithm can be interpreted as a procedure for finding critical points of a functional defined on the space of sections of a presheaf of probability space [19, 20]. Extension of this algorithm has been proposed for more general sheaves [21]. A sheaf-theoretic perspective on Belief Propagation has recently been proposed [22].

[1]Email: gregoire.sergeant-perthuis@sorbonne-universite.fr

BP is used for inference on Bayesian networks, on acyclic undirected graphical models, but also on loopy graphical models and factor graphs, for which many theoretical aspects remain to be understood [23, 24]. BP induces nonlinear dynamics on what are called "messages" but does not enjoy good functorial properties [22]. In this paper, we exploit the celebrated reformulation of fixed points of Generalized BP as critical points of the Bethe Free Energy to introduce a message-passing algorithm that is functorial and whose fixed points correspond to the critical points of the Bethe Free Energy and of Generalized Belief Propagation.

More precisely, the Belief Propagation algorithm is a message-passing procedure that enforces consistency conditions encoded by a presheaf over a partially ordered set, taking values in a finite set [6, 19, 20, 21, 7]. In particular, one can interpret BP and Generalized BP as dynamical systems on the Lagrange multipliers associated with the constraints satisfied by the sections of an associated presheaf, seeking beliefs that are compatible under marginalization.

Exploiting results from [20, 21, 7], we observe that BP and Generalized BP can be decomposed into a collection of operators of the form $\widetilde{\delta_F} \circ g_H^{\langle \cdot,\cdot \rangle} \circ \zeta_{F^\dagger} \circ d_{F^\dagger}$, the last of which breaks functoriality under natural transformations of presheaves. In our setting, natural transformations specify compatible ways of binning values of groups of variables, similar to a "local coarse graining".

By considering an alternative message-passing algorithm MP built from $\delta_F \circ g_H^{\langle \cdot,\cdot \rangle} \circ \zeta_{F^\dagger} \circ d_{F^\dagger}$ which computes exactly the same marginals that Generalized BP is designed to approximate, we obtain an algorithm that maps natural transformations $\phi : F \to G$ to linear maps, mimicking intertwining of linear operators, on the nonlinear message-passing operators $MP_F \to MP_G$ (see Theorem 1,3). We will explain all the operators appearing in the expressions of BP and MP in Section 2.2.

2 Preliminaries

Appendix A contains a brief overview of graphical models, factor graphs, and the motivation for introducing the Bethe free energy. In what follows, we consider the more general setting that we call *graphical presheaves*, introduced in [5], which extend graphical models and factor graphs, and then consider arbitrary presheaves as introduced in [21, 7].

Notations: A undirected graph will be denoted $G = (V, E)$, with V a finite set of vertices and E a collection of edges, i.e. subsets of I of cardinal 2. A hypergraph [25], denoted as $\mathscr{H} = (V, \mathscr{H}E)$, is defined by a finite set of vertices V and a collection of finite subsets of V called hyperedges. Both a graph and a hypergraph can be viewed as partially ordered sets, denoted respectively by $\mathscr{A}(G)$ and $\mathscr{A}(\mathscr{H})$, with the relation $v < e$ whenever a vertex $v \in V$ lies in the hyperedge $e \in E$. We consider a finite set E and probability distributions $p \in \mathbb{P}(E)$, where $\mathbb{P}(E)$ denotes the space of all probability distributions over E: for all $x \in E$, $p(x) \geq 0$ and $\sum_{x \in E} p(x) = 1$. We denote by $\mathbb{P}_{>0}(E)$ the space of strictly positive probability distributions, i.e., those p such that $p(x) > 0$ for all $x \in E$. We denote $E_V = \prod_{v \in V} E_v$ the space of configurations of atomic random variables $X_v \in E_v$, where each E_v is a finite set. Similarly, for $a \subseteq V$, we write $E_a = \prod_{v \in a} E_v$ for the product over indices in a. For $x_v \in E_v$, we let $x_a = (x_v)_{v \in a} \in E_a$ denote the corresponding projection; depending on context, we may also use x_a to denote a generic element of E_a.

2.1 Bethe Free Energy for factor graphs and extensions

Let $\mathscr{A}$ be a finite partially ordered set. The 'zeta-operator' of a poset $\mathscr{A}$, denoted ζ, is the operator from $\bigoplus_{a \in \mathscr{A}} \mathbb{R} \to \bigoplus_{a \in \mathscr{A}} \mathbb{R}$, where $\bigoplus_{a \in \mathscr{A}} \mathbb{R}$ is identified with the set of functions from $\mathscr{A}$ taking values in $\mathbb{R}$, defined as, for any $\lambda \in \bigoplus_{a \in \mathscr{A}} \mathbb{R}$ and any $a \in \mathscr{A}$, $\zeta(\lambda)(a) = \sum_{b \leq a} \lambda_b$. The zeta-operator of $\mathscr{A}$ is invertible (Proposition 2 [26]). We will call its inverse the Möbius inversion of $\mathscr{A}$, and we will denote it as μ. There is a collection $(\mu(a,b) \in \mathbb{Z}; b,a \in \mathscr{A}$ s.t. $b \leq a)$ such that, for any $\lambda \in \bigoplus_{a\in\mathscr{A}} \mathbb{R}$ and $a \in \mathscr{A}$, $\mu(\lambda)(a) = \sum\limits_{b\leq a} \mu(a,b)\lambda_b$. The coefficients $c(a)$, which appear in the inclusion-exclusion principle, are defined as $c(a) = \sum_{b\geq a} \mu(b,a)$. The reader can refer to Appendix C for a detailed presentation of Möbius inversion and the inclusion-exclusion principle. Let I be an index set of random variables, i.e., for each $i \in I$, there is a variable $X_i \in E_i$ taking values in a finite set E_i. Let $\mathscr{A}$ be a collection of subsets of I, i.e., $\mathscr{A} \subseteq \mathscr{P}(I)$ is a subset of $\mathscr{P}(I)$, the power set of I; in particular it is a poset for the inclusion relation, $\subseteq$, on subsets of I. Consider, for each $a \in \mathscr{A}$, a set of factors, $f_a : E_a = \prod_{i\in a} E_i \to \mathbb{R}$. Then the joint distribution $P \in \mathbb{P}(E_I)$ factors according to this collection of factors when, for any $x \in E_I$, $P(x) = \prod_{a\in\mathscr{A}} f_a(x_a)$. The Generalized Bethe free energy [5, 6] associated with $\mathscr{A}$ and the collection $(f_a; a \in \mathscr{A})$ is defined as, for any collection of probability distributions $Q = (Q_a \in \mathbb{P}(E_a); a \in \mathscr{A})$,

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

with, for any $a \in \mathscr{A}$, $c(a) = \sum_{b\geq a} \mu(b,a)$, and for $x = (x_i; i \in a)$ an element of E_a, $H_a(x_a) = \sum_{b\subseteq a} -\ln f_b(x_b)$, where $x_b = (x_i, i \in b)$. We recover the case of graphical models when $\mathscr{A} = \mathscr{A}(G)$ for a graph G and of factor graphs when $\mathscr{A} = \mathscr{A}(\mathscr{H})$ for a hypergraph $\mathscr{H}$. The collection of $Q_a, a \in \mathscr{A}$ that we are interested in are the ones compatible under marginalization, i.e., when $b \subseteq a$, we require that Q_b is the marginal of Q_a when summing out the variables $X_{a\setminus b}$. This constraint appears naturally when introducing presheaves that we shall define now.

Definition 1 (Graphical presheaves)**.** *Let I be a finite set, and let $\mathscr{A} \subseteq \mathscr{P}(I)$ be a sub-poset of the powerset of I. Let E_i, $i \in I$, be finite sets. For $a \in \mathscr{A}$, define $E_a := \prod_{i\in a} E_i$. Let $F_a := E_a$, and for $b \subseteq a$, let $F^a_b : E_a \to E_b$ be the projection map from $\prod_{i\in a} E_i$ to $\prod_{i\in b} E_i$. The functor F is called a graphical presheaf from $\mathscr{A}$ to finite sets.*

The probabilistic sections of a graphical presheaf are the collection of probability distributions $(Q_a \in \mathbb{P}(E_a), a \in \mathscr{A})$ such that, for any $b \subseteq a$,

$$\forall x \in E_b, \quad Q_b(x) = \sum_{y : F^a_b(y) = x} Q_a(y).$$

In the rest of the text, we shall rewrite the previous sum $Q_b(x) = \sum_{y\in F_a} 1[F^a_b(y) = x] Q_a(y)$, where $1[x \in A]$ is the indicator function which equals 0 if $x \notin A$ and 1 if $x \in A$.

The probabilistic sections of a graphical presheaf F are the sections, in the classical sense, of a presheaf that we will denote as $\mathbb{P}F$ and define just after. See Appendix B for a review of the classical definitions and constructions related to presheaves. Sections of presheaf R correspond to elements $v_a \in R(a)$, $a \in \mathscr{A}$ that satisfy the following compatibility property: whenever $b \leq a$, then $R^a_b(v_a) = v_b$. The set of sections of a presheaf R is denoted by $\lim R$. The presheaf $\mathbb{P}F$ associates to each element $a \in \mathscr{A}$ the space of probability distributions $\mathbb{P}(E_a)$, viewed as a finite-dimensional vector space, where E_a is the corresponding set. For inclusions $b \subseteq a$, it associates the marginalization map $\mathbb{P}F^a_b : \mathbb{P}(E_a) \to \mathbb{P}(E_b)$, which sends $P \in \mathbb{P}(E_a)$ to the pushforward measure defined as follows: for any $x \in E_b$, $\mathbb{P}F^a_b(P)(x) = \sum_{y\in E_a} 1[F^a_b(y) = x] P(y)$.

The optimization problem for variational inference graphical presheaves F, as introduced in [5, 6], involves minimizing the Bethe Free Energy $F_{\text{Bethe}}(Q_a; a \in \mathscr{A})$ under the convex constraint that the $(Q_a; a \in \mathscr{A})$ are probabilistic sections of $\mathbb{P}F$, i.e., $(Q_a; a \in \mathscr{A}) \in \lim \mathbb{P}F$.

Graphical presheaves have the drawback of being very restrictive: the space E_a must be a product of atomic sets E_i for $i \in I$, with $i \in a$, corresponding to an 'atomic' variable $X_i \in E_i$. Furthermore, the maps that relate values in E_a to E_b, whenever $b \subseteq a$, this must correspond to ignoring the variables $X_{a\setminus b}$ following the projections from $E_a = \prod_{i\in a} E_i$ to $E_b = \prod_{i\in b} E_i$. These variables X_a for $a \in \mathscr{A}$ correspond to the hidden variables that build the prior governing the dependencies among observed variables. It seems natural to rather associate with each $a \in \mathscr{A}$ a variable $Z_a \in F_a$ taking values in any finite set F_a. Furthermore, in this more general setting, one would want any possible relationship between Z_a and Z_b, i.e., any function from F_a to F_b whenever $b \subseteq a$. In [21, 7], a generalization of the Bethe free energy and of the Belief Propagation algorithm are proposed for arbitrary variables $Z_a \in F_a$ indexed over any partially ordered set $(\mathscr{A}, \leq)$. The relationships between Z_a and Z_b when $b \leq a$ are given by stochastic maps, i.e., stochastic matrices whose rows consist of positive real values that sum to one. For the development of our paper, we only need the more restrictive case where the maps are deterministic, i.e., we allow any function $F_b^a : F_a \to F_b$ to relate the variables Z_a and Z_b when $b \leq a$. For this reason, in what follows, our main object of interest are presheaves from a poset $\mathscr{A}$ taking values in finite sets and the associated presheaf of probability distributions and their extension to probability distribution and vector spaces. We will denote **FinSet** the category whose objects are finite sets and whose morphisms are maps between these sets (see Appendix B).

Definition 2 (Presheaf in finite sets and extensions)**.** *Let F be a presheaf over a finite poset $(\mathscr{A}, \leq)$ taking values in finite sets, i.e. a function that assigns to each element $a \in \mathscr{A}$ a finite set F_a and a collection of functions $F_b^a : F_a \to F_b$ for any pair $b, a \in \mathscr{A}$ such that $b \leq a$, satisfying the condition that whenever $c \leq b \leq a$, $F_c^b \circ F_b^a = F_c^a$.*

We call the probabilistic extension of F the presheaf $\mathbb{P}F$, which assigns to each element $a \in \mathscr{A}$ the set of probability distributions $\mathbb{P}(F_a)$. To any pair of elements $b, a \in \mathscr{A}$ such that $b \leq a$, it assigns the map $\mathbb{P}F_b^a : \mathbb{P}(F_a) \to \mathbb{P}(F_b)$, which sends a probability distribution $P \in \mathbb{P}(F_a)$ to its pushforward measure. For any $x \in F_b$, this is defined as

$$\mathbb{P}F_b^a(P)(x) = \sum_{y\in F_a} 1[F_b^a(y) = x]P(y).$$

Since F_a is a finite set, $\mathbb{P}(F_a) \subseteq \mathbb{R}^{F_a}$ is a simplex in the vector space of real-valued functions $\mathbb{R}^{F_a}$. Furthermore, $\mathbb{P}F_b^a$ is the restriction of a linear map from $\mathbb{R}^{F_a}$ to $\mathbb{R}^{F_b}$. We denote the associated linear map by $\widetilde{F}_b^a : \mathbb{R}^{F_a} \to \mathbb{R}^{F_b}$, $\widetilde{F}_b^a(\lambda)(x) = \sum_{y\in F_a} 1[F_b^a(y) = x]\lambda(y)$. The collection of spaces $(\widetilde{F}_a := \mathbb{R}^{F_a}; a \in \mathscr{A})$ and of linear maps $(\widetilde{F}_b^a : \mathbb{R}^{F_a} \to \mathbb{R}^{F_b}; a, b \in \mathscr{A}$ such that $b \leq a)$, defines a presheaf over $\mathscr{A}$ taking values in finite vector spaces; we shall not it as $\widetilde{F}$.

The dual map $\widetilde{F}_b^{a,*} : \widetilde{F}_b^* \to \widetilde{F}_a^*$ sends linear forms $l : \widetilde{F}_b \to \mathbb{R}$ to $\widetilde{F}_b^{a,*}(l) = l \circ \widetilde{F}_b^a$. The collection of vector spaces $(\widetilde{F}_a^*; a \in \mathscr{A})$ and linear maps $\widetilde{F}_b^{a,*}$ for $b, a \in \mathscr{A}, b \leq a$ defines a functor over $\mathscr{A}$ taking values in vector spaces. We shall denote this functor by $\widetilde{F}^*$. When a scalar product $\langle\cdot,\cdot\rangle_a$ is specified on each vector space $\widetilde{F}_a$, there is an isomorphism $f_{\langle\cdot,\cdot\rangle_a} : F_a^* \to F_a$ which allows us to identify $\widetilde{F}_b^{a,*}$ with its adjoint $\widetilde{F}_b^{a,\dagger}$, defined as $\widetilde{F}_b^{a,\dagger} = f_{\langle\cdot,\cdot\rangle_a} \circ \widetilde{F}_b^{a,*} \circ f^{-1}_{\langle\cdot,\cdot\rangle_a}$.

The Generalized Bethe Free Energy for presheaves F from a finite poset taking values in finite sets, and a collection of Hamiltonians $H_a : F_a \to \mathbb{R}$, has the same expression as Equation 2.1, where $c(a) = \sum_{b\geq a} \mu(b, a)$ are defined using the Möbius inversion of the poset $\mathscr{A}$, and Q are elements of $\lim F$.

2.2 The General Belief Propagation Algorithm for Factor Graphs, Graphical Presheaves, and Generalization to presheaves of Sets

The General Belief Propagation algorithm is used to find the critical points of the Generalized Bethe Free Energy. A classical result states that the fixed points of this algorithm correspond to the critical points of that free energy, which we state in Proposition 1 and reprove in Appendix D. Let us now recall the expression of this algorithm following [5, 6].

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. 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 $m_{a\to b} \in \mathbb{R}^{E_b}_{>0}$ and bottom-up messages $n_{b\to a} \in \mathbb{R}^{E_b}_{>0}$.

These messages are related as follows:

$$\forall a, b \in \mathscr{A}, \text{s.t. } b \subseteq a, \quad n^t_{b\to a} = \prod_{\substack{c: b\subseteq c\\ c \nsubseteq a}} m^t_{c\to b} \tag{2.2}$$

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

$$\forall a \in \mathscr{A}, \forall x_a \in E_a \quad b^t_a(x_a) \propto e^{-H_a(x_a)} \prod_{\substack{b\in\mathscr{A}:\\ b\subseteq a}} n^t_{b\to a}(x_b) \tag{2.3}$$

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,

$$\forall a, b \in \mathscr{A} \text{s.t.} b \subseteq a, \forall x \in E_b, \quad m^{t+1}_{a\to b}(x) = m^t_{a\to b}(x) \frac{\sum_{y\in E_a: y_b = x} b^t_a(y)}{b^t_b(x)} \tag{2.4}$$

One observes that in the previous Equation 2.4, 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 $b \leq a$,

$$m^{t+1}_{a\to b}(x) = m^t_{a\to b}(x) \frac{\sum_{y\in E_a: y_b = x} e^{-H_a(y)} \prod_{\substack{c\in\mathscr{A}:\\ c\subseteq a}} \prod_{\substack{d: c\subseteq d\\ d\nsubseteq a}} m^t_{d\to c}(y_c)}{e^{-H_b(x)} \prod_{\substack{c\in\mathscr{A}:\\ c\subseteq b}} \prod_{\substack{d: c\subseteq d\\ d\nsubseteq b}} m^t_{d\to c}(x_c)} \tag{2.5}$$

We denote the collection $(m_{a\to b}; a, b \in \mathscr{A}, b \leq a)$ as m. We denote $\mathrm{BP} : \prod_{a,b: b\subseteq a} \mathbb{R}^{E_b} \to \prod_{a,b: b\subseteq a} \mathbb{R}^{E_b}$ as the operator underlying the update rule of Equation 2.5, 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 2.3 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 $[\mathrm{BP}]([m]) = [\mathrm{BP}(m)]$.

Proposition 1 (Yedidia, Freeman, Weiss, Peltre)**.** *Let I be an finite set, and $\mathscr{A} \subseteq \mathscr{P}(I)$ a collection of subsets of I; let F be a graphical presheaf. Let $(m_{a\to b} \in \mathbb{R}^{E_b}_{>0}, a, b \in \mathscr{A}$ s.t. $b \subseteq a)$ be a fix point of the Generalized Belief Propagation up to a multiplicative constant, i.e. $[m] = [BP]([m])$. Let $(b_a, a \in \mathscr{A})$ be the associated beliefs normalized so that each $b_a \in \mathbb{P}(E_a)$ (Equation 2.3). Then $(b_a, a \in \mathscr{A})$ is a critical point of F_{Bethe} under the constraint that $p \in \lim \mathbb{P}F$. Furthermore, any critical point of F_{Bethe} in $\lim \mathbb{P}F$ is a belief associated to a fixed point of BP.*

Proof. See Theorem 5.15 in [6] or Theorem 5 in [5]. This result is also a corollary of Theorem 2.2. [21]. In Appendix D, we reprove the result.

□

An extension of the Belief Propagation algorithm to presheaves from a finite poset taking values in finite sets can be found in [21, 7]. In their work, for each presheaf F, they propose a message-passing algorithm, that we will denote as MP, whose set of fixed points corresponds to the critical points of the Generalized Bethe Free Energy. When considering graphical presheaves, these message-passing algorithms slightly differ from the Belief Propagation algorithm but have the same fixed points, see Appendix E for a description of the shared properties of the two algorithms. However, as we will show, MP behaves well under natural transformations $\phi : F \to G$, whereas BP does not. In Appendix F, we show, however, that the generalized Belief Propagation algorithm is not functorial. Let us now recall their message-passing algorithm, MP. To do so, following [21, 7], we need to introduce the elementary operators from which MP is built.

For a functor G from $\mathscr{A}$ to $\mathbb{R}$-vector spaces, we define μ_G as, for any $a \in \mathscr{A}$ and $v \in \bigoplus_{a\in\mathscr{A}} G_a$, $\mu_G(v)(a) = \sum_{b\leq a} \mu(a,b) G^b_a(v_b)$. Let F be a presheaf from a finite poset $\mathscr{A}$ to finite sets **FinSet**. Let us call $\mathrm{FE} : \prod_{a\in\mathscr{A}} \mathbb{R}^{E_a} \to \prod_{a\in\mathscr{A}} \mathbb{R}$ the extension of F_{Bethe} to real valued functions, i.e.

$$\forall a \in \mathscr{A}, \quad \mathrm{FE}(h)(a) = \sum_{x_a\in F_a} h_a(x_a) H_a(x_a) + \sum_{x_a\in F_a} h_a(x_a) \ln h_a(x_a) \tag{2.6}$$

Let us denote $b \leq a$ as $a \to b$. Let $\delta_F : \bigoplus_{a\in\mathscr{A}} F_a \to \bigoplus_{a,b\in\mathscr{A}:b\leq a} F_b$ be defined as, for $(v_a \in F_a, a \in \mathscr{A})$:

$$\forall a, b \in \mathscr{A}, \text{s.t. } b \leq a, \quad \delta_F(v)(a \to b) = F^a_b(v_a) - v_b \tag{2.7}$$

For a (covariant) functor G, $d_G : \bigoplus_{a,b\in\mathscr{A}:b\leq a} G(b) \to \bigoplus_{a\in\mathscr{A}} G_a$ is defined for $(v_{a\to b}, a, b \in \mathscr{A}$ such that $a \geq b)$ as

$$d_G(v)(a) = \sum_{b:b\leq a} G^b_a(v_{a\to b}) - \sum_{b:a\leq b} v_{b\to a}. \tag{2.8}$$

Here, G will be either F^* or $F^\dagger$ when each space F_a is equipped with a scalar product $\langle\cdot,\cdot\rangle_a$.

The ζ function of a functor G plays an import role in the Belief propagation algorithm. $\zeta_G : \bigoplus_{a\in\mathscr{A}} G_a \to \bigoplus_{a\in\mathscr{A}} G_a$ is defined as, for $v \in \bigoplus_{a\in\mathscr{A}} G_a$,

$$\zeta_G(v)(a) = \sum_{b \leq a} G_a^b(v_b) \tag{2.9}$$

Proposition 2. *Let F be a presheaf from a finite to finite sets. Let $H_a \in \mathbb{R}^{F_a}, a \in \mathscr{A}$, be a collection of Hamiltonians. The differential $dFE : \bigoplus_{a\in\mathscr{A}} \widetilde{F}_a \to \bigoplus_{a\in\mathscr{A}} \widetilde{F}_a^*$ that sends h to $d_h FE$ is invertible. We will denote its inverse by $g_H : \bigoplus_{a\in\mathscr{A}} \widetilde{F}_a \to \bigoplus_{a\in\mathscr{A}} \widetilde{F}_a$.*

Proof. Recall that $\mathrm{FE}_a(h_a) = \sum_{x_a} h_a(x_a)H_a(x_a) + \sum_{x_a} h_a(x_a)\ln h_a(x_a)$, where $h \in \mathbb{R}^{E_a}$, and d_x denotes the differential of the function at point x. Therefore,

$$\mathrm{d}_h\mathrm{FE}_a = \sum_{x_a} \mathrm{d}h_a(x_a)(H_a(x_a) + \ln h_a(x_a) + 1). \tag{2.10}$$

with $\mathrm{d}h_a(x_a) \in \bigoplus \widetilde{F}_a^*$ being the linear form acting as $(h_b(x_b); b \in \mathscr{A}, x_b \in F_b) \mapsto h_a(x_a)$. Consider the scalar products $\langle h_a, h_a^{'}\rangle_a = \sum_{x_a\in F_a} h_a(x_a)h_a^{'}(x_a)$ for each $a \in \mathscr{A}$; denote $y_a \in \widetilde{F}_a$ the identification of $d_h\mathrm{FE}_a \in F_a^*$ in F_a; then the reformulation of Equation 2.10 is the following,

$$\forall x_a \in F_a, y_a(x_a) = H_a(x_a) + \ln h_a(x_a) + 1$$

The previous equation is equivalent to

$$\forall x_a \in F_a, h_a(x_a) = e^{-H_a(x_a)+y_a(x_a)-1}$$

Therefore, we define

$$\forall a \in \mathscr{A}, \forall x_a \in E_a, \quad g^+_{H_a,a}(l)(x_a) = e^{-H_a(x_a)+l_a(x_a)-1}. \tag{2.11}$$

The function $g_H = (g_{H_a,a}; a \in \mathscr{A})$ is the inverse of dFE when the image of dFE, i.e., $\bigoplus_{a\in\mathscr{A}} \widetilde{F}_a^*$, is identified with $\bigoplus_{a\in\mathscr{A}} \widetilde{F}_a$ through the scalar product $\langle h, h_1\rangle = \sum_{a\in\mathscr{A}} \langle h_a, h_{1,a}\rangle_a$.

□

Remark 1. Any other choice of scalar product $\langle\cdot,\cdot\rangle_1$ on $\widetilde{F}_a$ induces another map $g_a^{\langle\cdot,\cdot\rangle_1} : \widetilde{F}_a \to \widetilde{F}_a$ defined as $g_a \circ f_{\langle\cdot,\cdot\rangle} \circ f^{-1}_{\langle\cdot,\cdot\rangle_1}$, where $\langle\cdot,\cdot\rangle$ is given by $\langle h, h_1\rangle = \sum_{x_a\in E_a} h(x_a)h_1(x_a)$. The way the inverse $g_H = \widetilde{F}_a^* \to \widetilde{F}_a$ of Proposition 2 is represented as a function $g_H^{\langle\cdot,\cdot\rangle}$ depends on the choice of scalar product on F_a. For the scalar products $\langle h_a, h_a^{'}\rangle = \sum_{x_a} h_a(x_a)h_a^{'}(x_a)$, we denote the identification $f^{\langle\cdot,\cdot\rangle} : \widetilde{F}_a^* \to \widetilde{F}_a$ as f_+ and the associated inverse map as g_H^+.

In what follows, we denote $\bigoplus_{a,b:b\leq a} F_b$ as $F_\to$ to simplify the notations.

Definition 3 (Message passing algorithms [21])**.** *Choose for each $a \in \mathscr{A}$ a scalar product $\langle\cdot,\cdot\rangle_a$ on $\widetilde{F}_a$; equip $F_\to$ with the following scalar product: $\langle l, l_1\rangle = \sum_{b,a:b\leq a}\langle l_{a\to b}, l_{1,a\to b}\rangle_b$. The message-passing algorithm $MP_{F,H}$ for inference on a presheaf F from a finite poset $\mathscr{A}$ to finite sets, with Hamiltonians $(H_a \in \mathbb{R}^{E_a}; a \in \mathscr{A})$, is defined as follows:*

$$\forall l \in F_\to, \quad MP_{F,H}(l) = l + \delta_F \circ g_H^{\langle\cdot,\cdot\rangle} \circ \zeta_{F^\dagger} \circ d_{F^\dagger}(l) \tag{2.12}$$

The algorithm starts with a random initialization of messages $l_0 \in F_\to$, and l^{t+1} is updated into the value $MP_{F,H}(l^t)$. We will denote $\Delta MP_{F,H}$ the increment $\delta_F \circ g_H \circ \zeta_{F^\dagger} \circ d_{F^\dagger}$.

Proposition 3. *Let F be a presheaf from a poset $\mathscr{A}$ taking values in finite sets. Let $(H_a \in \mathbb{R}^{E_a}; a \in \mathscr{A})$ be a collection of Hamiltonians. Let $(l_{a\to b} \in \widetilde{F}_b^*; a,b \in \mathscr{A} : b \leq a)$ be a fixed point of $MP_{F,H}$, i.e., $MP_{F,H}(l) = l$. Let $b_a \in \mathbb{P}(F_a)$ be the unique probability distribution such that*

$$b_a \propto g_H \circ \zeta_{F^*} \circ d_F(l).$$

Then $(b_a, a \in \mathscr{A})$ is a critical point of F_{Bethe} under the constraint that $p \in \lim \mathbb{P}F$. Furthermore, any critical point of F_{Bethe} in $\lim \mathbb{P}F$ is a belief associated to a fixed point of $MP_{F,H}$.

Proof. Theorem 2.2. [21]. In Appendix D, we reprove the result. □

Remark 2. Regardless of the choice of a scalar product, the fixed points of the associated message passing algorithm (Proposition 3) correspond to the same critical points. Indeed, the critical points are defined without the need to specify any scalar product. Furthermore, as explained in Appendix E, the fixed points of BP and MP are in correspondence; both algorithms (Propositions 1 and 3) allow us to characterize the critical points of the Bethe free energy.

3 Main results

3.1 Functoriality of ΔMP

Let $\mathscr{A}$ be a finite poset, and let F, G be two presheaves taking values in finite sets. A natural transformation $\phi : F \to G$ is a collection of maps $\phi_a : F_a \to G_a$ compatible with the maps induced by the relation $b \leq a$, i.e., $G_b^a \circ \phi_a = \phi_b \circ F_b^a$ whenever $b \leq a$ (see Appendix B).

Recall that the the presheaves we are concened with have as source a finite poset and take values in finite sets with function for each relation $b \leq a$. Each presheaf encodes a collection of local variables that interact via maps and specifies how beliefs on each local variable should be compatible through the marginalisation of probability distributions when one wants to form a global description of that collection of variables. A natural transformation then assigns to each element $x \in F_a$ an element $\phi_a(x) \in G_a$ in such a way that these assignments are compatible with the maps relating a collection of variables X_a to another X_b whenever $b \leq a$. A very simple example arises from factor graphs. Let $\mathscr{H} = (V, \mathscr{E})$ be the hypergraph associated to a factor graph. For each vertex $v \in V$, let $f_v \colon E_v \to B_v$ be any function; this function can be interpreted as sending each $x_v \in E_v$ to its a bin $f_v(x_v) \in B_v$. The graphical presheaf induced by the family $(E_v; v \in V)$ and the hypergraph $\mathscr{H}$ is defined by the collection of sets $E_a = \prod_{v\in e} E_v$ for each hyperedge $e \in \mathscr{E}$, together with the projection maps $E_e \to E_v$ for $v \in e$. Then the induced maps $f_a \colon \prod_{v\in a} E_v \to \prod_{v\in a} B_v$ are natural in $a \subseteq V$, and thus define a natural transformation between the graphical presheaf associated to $(E_v; v \in V)$ and the graphical presheaf associated to $(B_v; v \in V)$ over the same poset $\mathscr{A}(\mathscr{H})$.

A natural transformation from F to G induces a natural transformation between the presheaves $\mathbb{P}F$ to $\mathbb{P}G$ and $\widetilde{F}$ to $\widetilde{G}$ defined respectively as, for any $P \in \mathbb{P}(F_a)$ and $x \in G_a$, $\mathbb{P}\phi(P)(x) = \sum_{y\in F_a} 1[\phi_a(y) = x]P(y)$ and for any $h \in \mathbb{R}^{F_a}$, $\widetilde{\phi}(h)(x) = \sum_{y\in F_a} 1[\phi_a(y) = x]h(y)$. Similarly, there is a natural transformation of functors $\phi^* : \widetilde{G}^* \to \widetilde{F}^*$, i.e., a collection $\phi_a^* : \widetilde{G}_a^* \to \widetilde{F}_a^*$. It is defined as follows: for any $a \in \mathscr{A}$ and $l \in \widetilde{G}_a^*$, $\phi_a^*(l) = l \circ \phi_a$. Similarly, when the spaces are equipped with a scalar product, there is a natural transformation $\phi^\dagger : \widetilde{G}^\dagger \to \widetilde{F}^\dagger$.

$\widetilde{\phi}, \phi^*, \phi^\dagger$ are extended into maps on the sources and targets of the operators δ_F, g_H, ζ_{F^*}, and d_F. ϕ extend to $\phi^\oplus : \bigoplus_{a\in\mathscr{A}} \widetilde{F}_a \to \bigoplus_{a\in\mathscr{A}} \widetilde{G}_a$ where for $v \in \bigoplus_{a\in\mathscr{A}} \widetilde{F}_a$, $\phi^\oplus(v)_a = \widetilde{\phi}(v_a)$; we denote $\phi^\oplus$ simply as ϕ. Similarly $\phi^{\oplus,*} : \bigoplus_{a\in\mathscr{A}} \widetilde{G}_a^* \to \bigoplus_{a\in\mathscr{A}} \widetilde{F}_a^*$, where $\phi^*(l)_a = l_a \circ \phi_a$; for the choice of scalar product on

$\widetilde{F}_a$, $\phi^{\oplus,\dagger} = \bigoplus_{a\in\mathscr{A}} \phi_a^\dagger$. Similarly, ϕ extends to $\phi_1 : F_\to \to G_\to$, where $\phi(v)_{a\to b} = \widetilde{\phi}(v_{a\to b})$ for $v \in F_\to$. The map $\phi_1^* : \widetilde{G}^*_\to \to \widetilde{F}^*_\to$ is defined as $\bigoplus_{a,b:b\leq a} \phi_b^*$ and $\phi_1^\dagger = \bigoplus_{a,b:b\leq a} \phi_b^\dagger$.

Theorem 1. *Let F,G be two presheaves from a finite poset $\mathscr{A}$ taking values in finite sets* ***FinSet****; let $\phi : F \to G$ be a natural transformation. Then, for any collection of Hamiltonians $(H_a : F_a \to \mathbb{R}; a \in \mathscr{A})$, and any choice of scalar product on $\widetilde{F}_a$ and $\widetilde{G}_a$ with $a \in \mathscr{A}$,*

$$\Delta MP_{G,\widetilde{H}} = \phi_1 \circ \Delta MP_{F,H} \circ \phi_1^\dagger \tag{3.1}$$

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

$$\forall y \in G_a, \quad \widetilde{H}_a(x) = -\ln \sum_{x\in F_a:\phi_a(x)=y} e^{-H_a(x)} \tag{3.2}$$

with the convention that if $x \notin \operatorname{im}\phi_a$, then $\widetilde{H}_a = +\infty$.

Proof. In Appendix G, we give a detailed proof of Theorem 1. It relies on showing the following properties:

$$\begin{aligned} \phi^* \circ d_G &= d_F \circ \phi_1^* \\ \phi^* \circ \zeta_{G^*} &= \zeta_{F^*} \circ \phi^* \\ \phi_1 \circ \delta_F &= \delta_G \circ \phi \end{aligned} \tag{3.3}$$

One shows that for the choice of scalar product $\langle l, l_1\rangle_a = \sum_{x_a\in F_a} l(x_a) l_1(x_a)$, then,

$$\phi \circ g_H^+ \circ \phi^\dagger = g_{\widetilde{H}}^+$$

Recall that f_+ is the map that identifies F_a^* with F_a, and g_H^+ is the map associated with g_H by such an identification.

And one observes that for any other scalar product $\langle\cdot,\cdot\rangle$,

$$\phi \circ g_H^{\langle\cdot,\cdot\rangle} \circ \phi^\dagger_{\langle\cdot,\cdot\rangle} = \phi \circ g_H^+ \circ f_+ \circ f^{-1}_{\langle\cdot,\cdot\rangle} \circ \phi^\dagger_{\langle\cdot,\cdot\rangle}$$

as

$$g_H^{\langle\cdot,\cdot\rangle} = g_H^+ \circ f^+ \circ f^{-1}_{\langle\cdot,\cdot\rangle}.$$

Then,

$$g_H^{\langle\cdot,\cdot\rangle} \circ \phi^\dagger_{\langle\cdot,\cdot\rangle} = g_H^+ \circ f_+ \circ \phi^* \circ f^{-1}_{\langle\cdot,\cdot\rangle} = g_H^+ \circ f_+ \circ \phi^* \circ f_+^{-1} \circ f_+ \circ f^{-1}_{\langle\cdot,\cdot\rangle}.$$

As

$$\phi \circ g_H^+ \circ f_+ \circ \phi_* \circ f_+^{-1} = \phi \circ g_H^+ \circ \phi^\dagger = g_{\widetilde{H}}^+,$$

then,

$$\phi \circ g_H^{\langle\cdot,\cdot\rangle} \circ \phi^\dagger_{\langle\cdot,\cdot\rangle} = \phi \circ g_{\widetilde{H}}^+ \circ f_+ \circ f^{-1}_{\langle\cdot,\cdot\rangle}.$$

Therefore,

$$\phi \circ g_H^{\langle\cdot,\cdot\rangle} \circ \phi^\dagger_{\langle\cdot,\cdot\rangle} = g_{\widetilde{H}}^{\langle\cdot,\cdot\rangle}.$$

$\square$

Remark 3. By convention, we allow $H_a(x_a)$ to be equal to $+\infty$ for some $a \in \mathscr{A}$ and $x_a \in F_a$. If, furthermore, the collection of subsets $S_a = \{x_a \mid H_a(x_a) = +\infty\}$ is such that $\overline{S_a} = F_a \setminus S_a$ defines a subobject of F that we will denote as $\overline{S}$, i.e. if $F_b^a(\overline{S_a}) \subseteq \overline{S_b}$, then for any $l \in F_{\rightarrow}$,

$$\forall a, b, b \leq a, \quad \forall x_b \in S_b, \quad MP(l)_{a\to b}(x_b) = 0$$

The proof of these statements can be written straightforwardly or can be seen as a consequence of Theorem 1 for $\phi : \overline{S} \hookrightarrow F$, the inclusion of the subobject $(\overline{S_a}; a \in \mathscr{A})$ into F. In particular, for any natural transformation $\phi : F \to G$, the image $\operatorname{im}\phi = (\operatorname{im}\phi_a; a \in \mathscr{A})$ is a subobject of G. To avoid having to handle points $a \in \mathscr{A}$ and $x_a \notin \operatorname{im}\phi_a$ for which $\tilde{H}_a(x_a) = +\infty$, one can restrict MP to messages $l \in G_{\rightarrow}$ that satisfy $l_{a\to b}(x_b) = 0$ for all $x_b \notin \operatorname{im}\phi_b$.

Similarly, one can choose as a convention that $BP(m)_{a\to b}(x_b) = -\infty$ for $x_b \in S_b$; more precisely, in both case for MP and BP both conventions imply that the associated beliefs are 0:

$$b_a(x_a) = g_H^{\langle\cdot,\cdot\rangle} \circ \zeta_{F^\dagger} \circ d_{F^\dagger}(\ln m)(a, x_a) = 0,$$

Then for $x_b \in S_b$, we have

$$F_b^a(b_a)(x_b) = 0.$$

Proposition 4 (The category **MPA**). *Let us now define the category of message passing algorithms, which we will denote as **MPA**. Its objects are maps $D_0 : H \times S \to H$, where H is a Hilbert space and S is a topological space; those maps correspond to dynamics from a Hilbert space H to itself, indexed by a parameter space S.*

Consider a second object, $D_1 : H_1 \times S_1 \to H_1$. A morphism between two such objects, $\phi : D_0 \to D_1$, is a pair (ϕ_0, ϕ_1), where $\phi_0 : H \to H_1$ is a bounded linear map and $\phi_1 : S \to S_1$ is a continuous map, such that

$$\forall x \in H_1, \forall s \in S, \quad D_1(x, \phi_1(s)) = \phi_0 \circ D_0(\phi_0^\dagger(x), s).$$

The composition $\phi \circ \psi$ of two morphisms ϕ, ψ is defined as the composition of their maps $\phi_0 \circ \psi_0$ and $\phi_1 \circ \psi_1$.

Proof. $\phi_0 \circ \psi_0$ and $\phi_1 \circ \psi_1$ are respectively bounded linear and convex; the composition is indeed a morphism, $\phi \circ \psi : D_0 \to D_2$, from D_0 to D_2 as:

$$\forall x \in H_2, \forall s \in S, \quad D_2(x, (\phi_1 \circ \psi_1)(s)) = \phi_0 \circ D_1(\phi_0^\dagger(x), \phi_1(s)) = \phi_0 \circ \psi_0 \circ D_0(\psi_0^\dagger \circ \phi_0^\dagger(x), s).$$

and so,

$$\forall x \in H_2, \forall s \in S, \quad D_2(x, (\phi_1 \circ \psi_1)(s)) = \phi_0 \circ \psi_0 \circ D_0((\phi_0 \circ \psi_0)^\dagger x, s)$$

□

Let us denote the category of presheaves from a poset $\mathscr{A}$ to finite sets as $\hat{\mathscr{A}_f}$. The space $\prod_{a \in \mathscr{A}} \mathbb{R}^{F_a}$ is a finite-dimensional vector space and therefore a topological space for any norm defined on it. Consider, for example, the $\|\cdot\|_\infty$ norm, defined as $\|H\|_\infty = \sup_{a \in \mathscr{A}} \sup_{x_a \in F_a} |H_a(x_a)|$.

Theorem 2. *Let $\mathscr{A}$ be a finite poset; let $F : \mathscr{A} \to$ **FinSet** be a presheaf from $\mathscr{A}$ to finite sets. Recall that $F_{\to} = \bigoplus_{a,b\in\mathscr{A}:b\leq a} F_a$; equip each $\tilde{F}_a = \mathbb{R}^{F_a}$ with any scalar product $\langle\cdot,\cdot\rangle_a$ which in turn equips $F_{\to}$ with a scalar product. The map $\Delta MP(F) : F_{\to} \times \prod_{a\in\mathscr{A}} \mathbb{R}^{F_a} \to F_{\to}$ is defined as,*

$$\forall l \in F_{\to}, H \in \prod_{a\in\mathscr{A}} \mathbb{R}^{F_a}, \quad \Delta MP(F)(l,H) = \Delta MP_H(l)$$

To a natural transformation $\phi : F \to G$ between two presheaves of $\hat{\mathscr{A}_f}$, $\Delta MP(\phi) = (\phi,\psi)$ is a morphism between $\Delta MP(F)$ and $\Delta MP(G)$ in MPA, where

$$\forall a \in \mathscr{A}, y_a \in G_a, \quad \psi(H)_a(y_a) = -\ln \sum_{x_a\in F_a\,:\,\phi_a(x_a)=y_a} e^{-H_a(x_a)}$$

*Furthermore, ΔMP is a functor from $\hat{\mathscr{A}_f} \to$ **MPA**.*

Proof. It is a direct consequence of Theorem 1. See Appendix G for the full proof. □

3.2 Functoriality of MP

Let F be a presheaf of $\hat{\mathscr{A}_f}$, and recall that for $m \in F_{\to}$, $\mathrm{MP}_{F,H}(m) = m + \Delta\mathrm{MP}_{F,H}(m)$. Let G be a second presheaf of $\hat{\mathscr{A}_f}$, and let $\phi : F \to G$ be a natural transformation. Then, by Theorem 1, $\mathrm{MP}_{G,\tilde{H}}(m) = m + \phi \circ \Delta\mathrm{MP}_{F,H}(\phi^\dagger(m))$; however, $\phi\circ\phi^\dagger$ is not necessarily equal to the identity map, and therefore $\mathrm{MP}_{G,\tilde{H}}$ is, in general, different from $\mathrm{MP}_{G,\tilde{H}} \circ \phi^\dagger$. We will show in this section that for a good choice of scalar products on the F_a's, with $a \in \mathscr{A}$, $\mathrm{MP}_{F,H}$ is also a functor.

Lemma 1. *Let F,G be two presheaves from a poset $\mathscr{A}$ to a finite measurable set, let $\phi : F \to G$ be a natural transformation. Consider on $\widetilde{G}_a$ the standard scalar product $\langle h, j\rangle = \sum_{x\in G_a} h(x_a) j(x_a)$; for any $a \in \mathscr{A}$, consider the following scalar product on $\widetilde{F}_a$:*

$$\langle f,g\rangle_a^\phi = \sum_{x_a\in F_a} |\phi_a^{-1}(\{\phi_a(x_a)\})| f(x_a) g(x_a).$$

Then $\phi_a : \widetilde{F}_a \to \widetilde{G}_a$ and its adjoint $\phi^\dagger$ satisfy $\phi\circ\phi^\dagger(f) = f$ for any $f \in \prod_{a\in\mathscr{A}} \mathbb{R}^{\mathrm{im}\,\phi_a}$, i.e., $\phi\circ\phi^\dagger$ is the identity map on $\widetilde{\mathrm{im}\,\phi}$; furthermore, $\phi^\dagger$ is an isometry. In particular, if ϕ is a epimorphism, i.e. if for all $a \in \mathscr{A}$ ϕ_a is surjective, then $\phi\circ\phi^\dagger = id$.

Proof. Let $f \in \widetilde{G}_a$ and $g \in \widetilde{F}_a$, then,

$$\begin{aligned}\langle f, \phi(g)\rangle &= \sum_{y_a\in G_a} f(x_a) \sum_{x_a:\phi_a(x_a)=y_a} g(x_a) \\ &= \sum_{x_a\in F_a} f(x_a) g(\phi_a(x_a)) \frac{|\phi_a^{-1}(\{\phi_a(x_a)\})|}{|\phi_a^{-1}(\{\phi_a(x_a)\})|}\end{aligned}$$

Therefore, $\phi^\dagger(g)(x_a) = \frac{1}{|\phi_a^{-1}(\{\phi_a(x_a)\})|} g\circ\phi(x_a)$. Furthermore, when $\phi_a : F_a \to G_a$ is surjective,

$$\forall y_a \in G_a, \quad (\phi\circ\phi^\dagger g)(y_a) = \sum_{x_a:\phi(x_a)=y_a} \frac{1}{|\phi_a^{-1}(\{\phi_a(x_a)\})|} g\circ\phi(x_a).$$

When $\phi_a^{-1}(\{y_a\})$ is empty, the previous expression is 0. If it is not empty—which is the case when $\phi : F_a \to G_a$ is surjective—then

$$\sum_{x_a:\phi(x_a)=y_a} \frac{1}{|\phi_a^{-1}(\{\phi_a(x_a)\})|} g\circ\phi(x_a) = \sum_{x_a:\phi(x_a)=y_a} \frac{1}{|\phi_a^{-1}(\{y_a\})|} g\circ\phi(x_a) = g(y_a).$$

In particular for $f,g \in \tilde{\mathrm{im}}\,\phi_a$, one has that,

$$\langle \phi^\dagger(f), \phi^\dagger(g)\rangle_a^\phi = \langle \phi\circ\phi^\dagger f, g\rangle_a = \langle f,g\rangle_a,$$

which shows that $\phi^\dagger$ is an isometry.

□

Theorem 3. *Let $\mathscr{A}$ be a finite poset; let $F,G : \mathscr{A} \to$ **FinSet** be two presheaves from $\mathscr{A}$ to finite sets. Let $\phi : F \to G$ be a epimorphisms, i.e. a natural transformation made of surjective maps. Recall that $F_\to = \bigoplus_{a,b\in\mathscr{A}:b\le a} F_a$; equip each $\widetilde{F}_a = \mathbb{R}^{F_a}$ with any scalar product $\langle\cdot,\cdot\rangle_a^\phi$, which in turn equips $F_\to$ with a scalar product. Let for $l \in F_\to$ and $H \in \prod_{a\in\mathscr{A}} \mathbb{R}^{F_a}$, $MP(F)(l,H) = l + \delta_F \circ g_H^{\langle\cdot,\cdot\rangle^\phi} \circ \zeta_{F^\dagger} \circ d_{F^\dagger}(l)$; let $MP(\phi) = (\phi,\psi)$ with,*

$$\forall a\in\mathscr{A},\, y_a \in G_a, \quad \psi(H)_a(y_a) = -\ln \sum_{x_a\in F_a:\phi_a(x_a)=y_a} e^{-H_a(x_a)}.$$

Then MP is a functor from $\hat{\mathscr{A}_f}$ to ***MPA***.

Proof. Theorem 3 is a direct consequence of Lemma 1 and Theorem 2 as for any $l \in G_\to$,

$$\phi\circ \mathrm{MP}_{F,H}\circ\phi^\dagger(l) = \phi\circ\phi^\dagger(l) + \phi\circ\mathrm{MP}_{F,H}\circ\phi^\dagger(l) = \mathrm{MP}_{G,\widetilde{H}}(l)$$

For a detailed proof of the result, see Appendix G.

□

4 Acknowledgment

The authors would like to thank the PRMO Young FMJH Program for supporting this research.

References

[1] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, 1988.

[2] Kevin Gimpel and Daniel Rudoy. *Statistical inference in graphical models*. Massachusetts Institute of Technology, Lincoln Laboratory, 2008.

[3] 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.

[4] Marc Mezard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, Inc., USA, 2009.

[5] 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.

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

[7] Grégoire Sergeant-Perthuis and Nils Ruet. Inference on diagrams in the category of markov kernels (extended abstract). 2024. Presented at ACT 7.

[8] Toby St Clere Smithe. Copy-composition for probabilistic graphical models. *Electronic Proceedings in Theoretical Computer Science*, 429:146–173, September 2025.

[9] Dylan Braithwaite, Jules Hedges, and Toby St Clere Smithe. The compositional structure of bayesian inference. *arXiv preprint arXiv:2305.06112*, 2023.

[10] Toby St Clere Smithe and Marco Perin. Autobayes: A compositional framework for generalized variational inference, 2025.

[11] Sean Tull, Johannes Kleiner, and Toby St Clere Smithe. Active inference in string diagrams: A categorical account of predictive processing and free energy, 2023.

[12] Bart Jacobs, Márk Széles, and Dario Stein. Compositional inference for bayesian networks and causality. *Electronic Notes in Theoretical Informatics and Computer Science*, Volume 5-Proceedings of..., December 2025.

[13] Masaki Kashiwara and Pierre Schapira. Persistent homology and microlocal sheaf theory. *Journal of Applied and Computational Topology*, 2018.

[14] Robert Ghrist. Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.

[15] Jakob Hansen and Thomas Gebhart. Sheaf neural networks, 2020.

[16] Justin Curry. *Sheaves, cosheaves and applications*. PhD thesis, The University of Pennsylvania, 2013. arXiv:1303.3255.

[17] Federico Barbero, Cristian Bodnar, Haitz Sáez de Ocáriz Borde, Michael Bronstein, Petar Veličković, and Pietro Liò. Sheaf neural networks with connection laplacians. In *Topological, Algebraic and Geometric Learning Workshops 2022*, pages 28–36. PMLR, 2022.

[18] T. Papamarkou, T. Birdal, M. Bronstein, G. Carlsson, J. Curry, Y. Gao, M. Hajij, R. Kwitt, P. Liò, P. Di Lorenzo, V. Maroulas, N. Miolane, F. Nasrin, K. N. Ramamurthy, B. Rieck, S. Scardapane, M. T. Schaub, P. Veličković, B. Wang, Y. Wang, G. W. Wei, and G. Zamzmi. Position: Topological deep learning is the new frontier for relational learning. In *Proceedings of Machine Learning Research*, volume 235, pages 39529–39555, Jul 2024.

[19] Olivier Peltre. A homological approach to belief propagation and Bethe approximations. In *International Conference on Geometric Science of Information*, pages 218–227. Springer, 2019.

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

[21] Grégoire Sergeant-Perthuis. Regionalized optimization, 2022.

[22] Enrique ter Horst, Sridhar Mahadevan, and Juan Diego Zambrano. Categorical belief propagation: Sheaf-theoretic inference via descent and holonomy. *arXiv preprint arXiv:2601.04456*, 2026.

[23] 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.

[24] 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, 2022.

[25] Alain Bretto. *Hypergraph Theory: An Introduction*. Springer Publishing Company, Incorporated, 2013.

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

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

[28] Steffen L. Lauritzen. Lectures on graphical models. Lecture Notes, 2019.

[29] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[30] Terry P. Speed. A note on nearest-neighbour gibbs and markov probabilities. *Sankhyā: The Indian Journal of Statistics, Series A*, 1979.

[31] Jonathan S Yedidia, William T Freeman, and Yair Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on information theory*, 51(7):2282–2312, 2005.

[32] Grégoire Sergeant-Perthuis and Nils Ruet. Inference on diagrams in the category of Markov kernels (Extended abstract). In *7th International Conference on Applied Category Theory (ACT 7)*, Oxford (UK), United Kingdom, June 2024. David Jaz Myers and Michael Johnso.

[33] Saunders Mac Lane. *Categories for the working mathematician*, volume 5. Springer Science & Business Media, 2013.

[34] Emily Riehl. *Category theory in context*. Courier Dover Publications, 2017.

[35] Jakob Hansen and Robert Ghrist. Distributed optimization with sheaf homological constraints. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 565–571, 2019.

[36] Erik Carlsson, Gunnar Carlsson, and Vin De Silva. An algebraic topological method for feature identification. *International Journal of Computational Geometry & Applications*, 16(04):291–314, 2006.

[37] Pierre Baudot and Daniel Bennequin. The homological nature of entropy. *Entropy*, 17(5):3253–3318, 2015.

[38] Juan Pablo Vigneaux. Information structures and their cohomology. *Theory and Applications of Categories*, 35(38):1476–1529, 2020.

A Graphical Models, Hypergraphs and generalizations

A.1 Factor Graphs

Let $G = (V, E)$ be an undirected graph, where V is a finite set of vertices and E is a collection of undirected edges, i.e., subsets of V of cardinality 2. For $v \in V$, let us denote by ∂v the neighbors of v, i.e., the elements $v_1 \neq v$ such that $\{v_1, v\} \in E$. A subset a of V such that every two distinct vertices in the subset a are neighbors is called a clique; we denote the set of cliques of G by $\mathscr{C}(G)$. Associate to each vertex $v \in V$ a unique random variable X_v taking values in a finite set denoted by E_v. For a subset of vertices $a \subseteq V$, we denote by X_a the associated joint variable $(X_v; v \in a)$, by x_a a configuration of X_a, and by E_a the associated configuration space for X_a, i.e., $E_a = \prod_{v \in a} E_v$. Following these conventions, as an example, $x_a \in E_a$ and $E_V = \prod_{v \in V} E_v$. Let us denote the space of joint probability distributions of the collection of variables $(X_v, v \in V)$, as $\mathbb{P}(E_V)$; $\mathbb{P}_{>0}(E_V)$ will correspond to the space of strictly positive probability distributions, i.e. for any $x \in E_V$, $P(x) > 0$. An *undirected graphical model*, also known as a *Random Markov field* [27, 28, 29], with respect to G, is a joint probability distribution $P \in \mathbb{P}(E_V)$ such that for any vertex $v \in V$, the random variable X_v is conditionally independent of $X_{V \setminus \{v \cup \partial v\}}$ given $X_{\partial v}$, denoted as $X_v \perp\!\!\!\perp X_{V \setminus (v \cup \partial v)} \mid X_{\partial v}$. The Hammersley-Clifford theorem [27] asserts that, for a strictly positive joint probability distribution $P \in \mathbb{P}_{>0}(E_V)$, the joint distribution of a undirected graphical model factors on the cliques of the graphs, i.e. that there is a collection of real valued functions $(f_a \in \mathbb{R}_{>0}^{E_a}, a \in \mathscr{C}(G))$ such that, $\forall x \in E_V, P(x) = \prod_{a \in \mathscr{C}(G)} f_a(x_a)$.

Factor graphs are an extension of undirected graphical models, starting with a *hypergraph* rather than a graph. A hypergraph, denoted as $\mathscr{H} = (V, \mathscr{H}E)$, is defined by a finite set of vertices V and a collection of finite subsets of V called hyperedges [25]. Undirected Graphs are particular cases of hypergraphs, where each $a \in \mathscr{H}E$ has cardinality 2. Hypergraphs are themselves particular instances of *partially ordered sets*, which are sets equipped with a binary relation $\leq$ that is transitive (i.e., if $c \leq b$ and $b \leq a$, then $c \leq a$) relfexive ($x \leq x$) and antisymmetric (i.e., if $b \leq a$ and $a \leq b$, then $a = b$); for hypergraphs, the order relation is given by the membership relation: $v \leq a$ whenever $v \in a$, with $v \in V$ and $a \in \mathscr{H}E$. We denote the poset associated with a hypergraph as $\mathscr{A}(\mathscr{H})$, and the poset associated with a graph, seen as a hypergraph, as $\mathscr{A}(G)$. A factor graph [3, 4] is a hypergraph with a unique set of variables X_v, for $v \in V$, one per node $v \in V$ (taking values in a finite set E_v), and a set of *factors* $f_a : E_a \to \mathbb{R}$, one for each hyperedge $a \in \mathscr{H}E$. A joint distribution $P \in \mathbb{P}(E_V)$ associated with a factor graph satisfies the following decomposition: for any $x \in E_V$, as $P(x) = \prod_{a \in \mathscr{H}E} f_a(x_a)$.

A.2 Bethe Free Energy for graphical models

An undirected graphical model with respect to an acyclic graph, whose joint distribution $P \in \mathbb{P}_{>0}(E_V)$ is strictly positive (i.e., $P(x) > 0$ for $x \in E_v$), also factors according to its marginal distributions on the edges and vertices [30, 5]. More precisely, consider an acyclic undirected graph $G = (V, E)$, i.e., there are no paths of distinct edges starting at a vertex and ending at the same vertex. Let $P_a \in \mathbb{P}_{>0}(E_a)$ denote the marginal distribution of X_a; recall that it is defined as $P_a(z_a) = \sum_{y_{\overline{a}} \in E_{\overline{a}}} P(z_a, y_{\overline{a}})$, where $z_a \in E_a$, and $\overline{a} = V \setminus a$ is the complement of a. Let $d(v)$ denote the degree of node $v \in V$, i.e., the cardinality of ∂v. Then, one shows by induction on the number of vercices, $|V|$, that for any $x \in E_V$,

$$P(x) = \frac{\prod_{e \in E} P_e(x_e)}{\prod_{v \in V} P_v^{d(v)-1}}. \tag{A.1}$$

Recall that the entropy of a probability distribution $P \in \mathbb{P}(E_V)$ is defined as $S(P) = -\sum_{x \in E_V} P(x) \ln P(x)$. The factorization of Equation A.1 allows for rewriting the entropy of the joint distribution $S(P) = -\sum_{x \in E_V} P(x) \ln P(x)$ as

$$S(P) = \sum_{e \in E} S(P_e) - \sum_{v \in V} (d(v) - 1) S(P_v). \tag{A.2}$$

The previous decomposition of entropy for undirected graphical models on acyclic graphs (Equation A.2) motivates the introduction of the Bethe free energy, defined in the next paragraph as an approximate variational inference method for models with loops. Recall that in variational inference, one observes that for a joint distribution $P_{X,Y}$ over two variables $X \in E_X$ (the hidden variable) and $Y \in E_Y$ (the observed variable), the conditional probability $P_{X|Y}(x|y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}$ given the observation $Y = y$ is the distribution $Q^* \in \mathbb{P}(E_X)$ that minimizes the free energy defined as $F(Q) = -\sum_{x \in E_X} Q(x) \ln P_{X,Y}(x,y) - S(Q)$. In particular, the log-marginal distribution on the observed variables $-\ln P_Y(y) = -\ln \sum_{x \in E_X} P_{X,Y}(x,y)$ corresponds to the optimal value of $F(Q)$:

$$-\ln P_Y(y) = \inf_{Q \in \mathbb{P}(E_X)} F(Q)$$

The Bethe free energy exploits the re-expression of entropy, Equation A.2, to express the free energy solely in terms of local marginals $Q_e \in \mathbb{P}_{>0}(E_e)$ for $e \in E$ and $Q_v \in \mathbb{P}_{>0}(E_v)$ for $v \in V$; for an undirected graphical model $P \in \mathbb{P}_{>0}(E_V)$ with respect to an acyclic graph $G = (V, E)$, the *Bethe Free Energy*, denoted F_{Bethe}, is defined on the local marginals $(Q_e, Q_v; v \in V, e \in E)$ by the following expression:

$$F_{\text{Bethe}}(Q_a; a \in \mathscr{A}(G)) = \sum_{a \in \mathscr{A}(G)} c(a) \left[S(Q_a) - \mathbb{E}_{Q_a}[H_a] \right] \tag{A.3}$$

with the following correspondence: $c(v) = -(d(v) - 1)$ for nodes, $c(e) = 1$ for edges, and $H_e = -\ln P_e$, $H_v = 0$. The expression of the Bethe free energy is also well-defined for undirected graphical models with graphs containing cycles; in this case, $F_{\text{Bethe}}(Q_a, a \in \mathscr{A}(G))$ differs from $F(Q)$ but is expected to be an approximation of $F(Q)$.

Inference on undirected graphical models involves determining the infimum of $F_{\text{Bethe}}(Q_a; a \in \mathscr{A}(G))$, with $Q_a \in \mathbb{R}^{E_a}$ any collection of functions, the constraints that for any $a \in \mathscr{A}(G)$, Q_a is a strictly probability distribution over E_a, i.e. $\sum_{x_a \in E_a} Q_a(x_a) = 1$ and for $x_a \in E_a$ $Q_a(x_a) > 0$, but also a consistency constraint:

$$\forall e = \{v, v_1\} \in E, \forall x_v \in E_v, \quad Q_v(x_v) = \sum_{y \in E_{v_1}} Q_e(x_v, y)$$

This condition accounts for the fact that if there is a joint distribution $P \in \mathbb{P}(E_V)$ such that Q_e is the marginal distribution of P over X_e, i.e., $Q_e = P_e$, and similarly for vertices $Q_v = P_v$, then when $v \in e$, Q_e must be related to Q_v by marginalization.

More interestingly, following [31][5][6] such inference and the Bethe Free Energy extend to factor graphs and any collection of subsets of an index set of variables I. Their standard formulation is the basis of an extension of variational inference to presheaves from a poset to finite sets [21][32] which offers a framework general enough for us to then apply topological transformations to graphical models and trace their impact on the critical points of the associated Bethe Free Energies. In the next subsection we present both extensions; they rely on the inclusion-exclusion principle [26], which introduces weights $c(a)$ for subsets a of an index set I. These weights $c(a)$ account for counting the same set multiple times due to overlaps and have a natural formulation for functions over a partially ordered set taking values in $\mathbb{Z}$.

B Categories, posets, functors and natural transformations in data science

For an introduction to categories, see [33, 34]. A category $\mathbf{C}$ is a collection of objects $\mathscr{O}(\mathbf{C})$ and morphisms $\phi : A \to B$ between two objects. Two morphisms $\phi : A \to B$, $\phi_1 : B \to C$, whose target and domain are compatible, compose into a morphism $\phi_1 \circ \phi : A \to C$. The composition is associative, $\phi_2 \circ (\phi_1 \circ \phi) = (\phi_2 \circ \phi_1) \circ \phi$. We will denote categories in bold. A (covariant) functor $F : \mathbf{C} \to \mathbf{C}_1$ between two categroies $\mathbf{C}$ and $\mathbf{C}_1$ send objects A of $\mathscr{O}(\mathbf{C})$ to objects $F(A)$ of $\mathscr{O}(\mathbf{C}_1)$; it sends a morphism $\phi : A \to B$ to a morphism $F(\phi) : F(A) \to F(B)$ such that $F(\phi_1 \circ \phi) = F(\phi_1) \circ F(\phi)$. A contravariant functor sends morphisms ϕ to $F(\phi) : F(B) \to F(A)$ and satisfies $F(\phi_1 \circ \phi) = F(\phi) \circ F(\phi_1)$. A covariant functor F from $\mathbf{C}$ to $\mathbf{C}_1$ is denoted as $F : \mathbf{C} \to \mathbf{C}_1$, and a contravariant functor as $F : \mathbf{C}^{op} \to \mathbf{C}_1$.

The categories of sets and finite sets, denoted respectively by **Set** and **FinSet**, have as objects sets and finite sets, respectively, and as morphisms between two sets X, Y, functions from X to Y. A contravariant functor from a category $\mathbf{C}$ to **Set** is called a presheaf. In particular, contravariant functors from $\mathbf{C}$ to **FinSet** are presheaves. We will also encounter implicitly the categroy of finite vector spaces, which has as objects finite vector spaces and as morphims linear maps between two such vector space.

A natural transformation between two covariant or contravariant functors $F, G : \mathbf{C} \to \mathbf{C}_1$ or $F, G : \mathbf{C}^{op} \to \mathbf{C}_1$ is a collection of morphisms $\phi_A : F(A) \to G(A)$ indexed by objects A of $\mathscr{O}(\mathbf{C})$ such that for any morphism $\psi : A \to B$, we have $\phi_B \circ F(\psi) = G(\psi) \circ \phi_A$ when F and G are covariant functors, and $\phi_A \circ G(\psi) = F(\psi) \circ \phi_B$. A natural transformation ϕ between two covariant or contravariant functors $F, G : \mathbf{C} \to \mathbf{Set}$ is called a monomorphism if each ϕ_A is injective, and an epimorphism if each ϕ_A is surjective.

A partially ordered set (poset), denoted as $\mathscr{A}$, is a set equipped with a binary relation $\mathscr{R} \subseteq \mathscr{A} \times \mathscr{A}$ which satisfies:

1. Reflexivity: $\forall x \in \mathscr{A}, \quad x\mathscr{R}x$.
2. Transitivity: $\forall x,y,z \in \mathscr{A}$, if $x\mathscr{R}y$ and $y\mathscr{R}z$, then $x\mathscr{R}z$.
3. Antisymmetry: $\forall x,y \in \mathscr{A}$, if $x\mathscr{R}y$ and $y\mathscr{R}x$, then $x = y$.

$\mathscr{R}$ is usually dentoted as $\leq$. One associates to a poset a category $\mathscr{C}(\mathscr{A})$, whose objects $\mathscr{O}(\mathscr{A})$ are the elements of $\mathscr{A}$, and there is a unique morphism $\phi : b \to a$ between $a,b \in \mathscr{A}$ when $b \leq a$. We will denote the associated category $\mathscr{C}(\mathscr{A})$ simply as $\mathscr{A}$. A contravariant functor $F : \mathscr{A} \to \mathbf{C}$ from the poset $\mathscr{A}$ seen as a category to a category $\mathbf{C}$ is encoded into a collection of objects $(F_a, a \in \mathscr{A})$ and a collection of morphisms $(F^a_b : F_a \to F_b; a,b \in \mathscr{A}, b \leq a)$.

The set of sections of a presheaf F, from a finite poset $\mathscr{A}$ to **Set**, also called limit and denoted $\lim F$, is the set of collections of elements $x = (x_a \in F_a \mid a \in \mathscr{A})$ such that they are pairwise compatible: for any $a,b \in \mathscr{A}$ with $b \leq a$, $F^a_b(x_a) = x_b$.

The application of presheaves and sheaves in data science has gained more importance in recent years; they were introduced in the context of decentralized optimization, for deep learning architectures that enable heterogeneous descriptions of data (Sheaf neural networks) [16, 35, 15, 17]. These approaches involve introducing sheaves over cell complexes as a data structure [16] and using a Laplacian as the foundational building block for neural network layers. A generalization of inference on graphical models, which involves presheaves from a poset to the category of Markov kernels, was introduced in [7], and for decentralized optimization in [21]. They also appear in topological data analysis [36, 14, 13] and for information theory [37, 38].

C Möbius inversion of a poset and of a functor

The definition of the Bethe Free Energy relies on weights given by the inclusion-exclusion formula of a poset; we shall now recall its definition.

Definition 4 (Zeta operator of a poset)**.** *Let $\mathscr{A}$ be a finite poset. We call the 'zeta-operator' of a poset $\mathscr{A}$, denoted ζ, the operator from $\bigoplus_{a\in\mathscr{A}} \mathbb{R} \to \bigoplus_{a\in\mathscr{A}} \mathbb{R}$ defined as, for any $\lambda \in \bigoplus_{a\in\mathscr{A}} \mathbb{R}$ and any $a \in \mathscr{A}$,*

$$\zeta(\lambda)(a) = \sum_{b\leq a} \lambda_b \tag{C.1}$$

Proposition 5 (Reformulation of Proposition 2 [26], Rota' 64)**.** *Let $\mathscr{A}$ be a finite poset. The zeta-operator of $\mathscr{A}$ is invertible. We will call its inverse the Möbius inversion of $\mathscr{A}$, denoted μ. Furthermore, there is a collection $(\mu(a,b); b,a \in \mathscr{A}$ s.t. $b \leq a)$ such that, for any $\lambda \in \bigoplus_{a\in\mathscr{A}} \mathbb{R}$ and $a \in \mathscr{A}$,*

$$\mu(\lambda)(a) = \sum_{b\leq a} \mu(a,b)\lambda_b \tag{C.2}$$

We call the coefficient $(\mu(a,b), b, a$ s.t. $b \leq a)$ the Möbius coefficients of $\mathscr{A}$. In particular Proposition 5 implies that, for any $b,a \in \mathscr{A}$ such that $b \leq a$,

$$\sum_{c: b\leq c\leq a} \mu(a,c) = 1[b=a] \tag{C.3}$$

$$\sum_{c: b\leq c\leq a} \mu(c,b) = 1[b=a] \tag{C.4}$$

For a collection of values $\lambda_a \in \mathbb{R}$, $a \in \mathscr{A}$, we call the following expression $\sum_{a\in\mathscr{A}} \sum_{b\leq a} \mu(a,b)\lambda_b$ the inclusion-exclusion formula over a poset $\mathscr{A}$; this formula corresponds to the value one would attribute to a maximal element, denoted 1, added to $\mathscr{A}$:

$$\lambda_1 := \sum_{a\in\mathscr{A}} \sum_{b\leq a} \mu(a,b)\lambda_b \tag{C.5}$$

which can be rewritten as $\lambda_1 = \sum_{a\in\mathscr{A}} [\sum_{b\geq a} \mu(b,a)]\lambda_a$. We will denote $c(a) = \sum_{b\leq a} \mu(a,b)$ the weighted coefficients.

To find the classical inclusion-exclusion formula, consider I a finite set; the poset $\mathscr{A}$ is $(\mathscr{P}(I), \supseteq)$ with the reversed order. The quantities λ_a, for $a \in \mathscr{A}$, are $|A_i|$ when $a = i \in I$, and represent the cardinality of the sets A_i, as well as the cardinality of all possible intersections $|\cap_{i\in a} A_i|$ when $a \subseteq I$. In this setting, the maximal element 1 has a value $\lambda_1 = |\cup_{i\in I} A_i|$.

Let G be a (covariant) functor from a finite poset $\mathscr{A}$ to finite vector spaces. Define,

$$\forall v \in \bigoplus G_a, \quad \zeta_G(v)_a = \sum_{b\leq a} G_a^b(v_b)$$

and,

$$\forall v \in \bigoplus G_a, \quad \mu_G(v)_a = \sum_{b\leq a} \mu(a,b) G_a^b(v_b).$$

Then,

$$\eta_G \circ \mu_G(v)_a = \sum_{c,b:c\leq b\leq a} \mu(b,c) G_a^c(v_c) = v_a,$$

and similarly, $\mu_G \circ \eta_G = \mathrm{id}$.

D Proof of Proposition 1 and 3 already stated in the literature

Lemma 2 (Particular case of Theorem 2.1 [21]). *Let F be a presheaf from a finite poset $\mathscr{A}$ to the category of finite sets **FinSet**. A point $v \in \lim \widetilde{F}$ is a critical point of F_{Bethe} if and only if*

$$\left[\mu_{F^*} d_v FE\right]\big|_{\lim \widetilde{F}} = 0, \tag{D.1}$$

where $|_{\lim \widetilde{F}}$ means restricted to the vector space $\lim \widetilde{F}$.

The previous Equation (D.1) *is equivalent to the existence of $l \in F_{\rightarrow}$ such that*

$$\mu_{F^*} d_v FE = d_{F^*}(l). \tag{D.2}$$

Proof. One shows that for any $v \in \bigoplus_{a\in\mathscr{A}} \widetilde{F}_a$,

$$\mathrm{d}_v F_{\mathrm{Bethe}}\big|_{\lim \widetilde{F}} = \left[\mu_{F^*} d_v \mathrm{FE}\right]\big|_{\lim \widetilde{F}},$$

where

$$\mathrm{FE}_a(h_a) = \sum_{x_a} h_a(x_a) H_a(x_a) + \sum_{x_a} h_a(x_a) \ln h_a(x_a),$$

for $h \in \mathbb{R}^{E_a}$. Its differential is given by Equation (2.10). Furthermore, a parametrization of $\lim \widetilde{F}$ is given by $\operatorname{im} \mathrm{d}_{F^*}$. $\square$

As ζ_{F^*} is the inverse of μ_{F^*}, Equation (D.2) is equivalent to the existence of $l \in F_{\rightarrow}$ and $v \in \lim \widetilde{F}$ such that

$$d_v \mathrm{FE} = \zeta_{F^*} d_{F^*}(l).$$

D.1 Proof of Proposition 1

The belief propagation algorithm Equation 2.5 can be rewritten as,

$$\ln m^{t+1}_{a\to b} = \ln m^{t}_{a\to b} + \ln \widetilde{F}^a_b \left(g_H \circ \zeta_{F^\dagger} \circ d_{F^\dagger}(\ln m^t)_a\right) - \ln \left(g_H \circ \zeta_{F^\dagger} \circ \mathrm{d}_{F^\dagger}(\ln m^t)_b\right)$$

Let us denote $\Delta\mathrm{BP}(\ln m)_{a\to b} = \ln \widetilde{F}^a_b (g_H \circ \zeta_{F^\dagger} \circ d_{F^\dagger}(\ln m)_a) - \ln (g_H \circ \zeta_{F^\dagger} \circ \mathrm{d}_{F^\dagger}(\ln m)_b)$.

Consider the collection $(\ln m_{a\to b} + C_{a\to b}; a,b \in \mathscr{A} : b \leq a)$, where $C_{a\to b}$ is a constant, i.e. it does not depend on $x_b \in F_b$. Then there is a collection of constants $(C^{'}_{a\to b}; a,b \in \mathscr{A} : b \leq a)$ such that $\Delta\mathrm{BP}(\ln m_{a\to b} + C_{a\to b}) = \Delta\mathrm{BP}(\ln m_{a\to b}) + C^{'}_{a\to b}$. Therefore BP is a algorithm that preserves the equivalence classes $\{m + C\}$, we will denote that class as $[m]$ and $\overline{\Delta\mathrm{BP}}([m]) = [\Delta\mathrm{BP}(m)]$.

Let $[\ln m]^*$ be a fix point of the Belief Propagation algorithm, and let $\ln m^r$ be a representant of $[\ln m]^*$ i.e. $[\ln m]^* = [\ln m^r]$. There is a a collection of constants $(C_{a\to b}; a,b \in \mathscr{A} : b \leq a)$,

$$C_{a\to b} = \ln \widetilde{F}^a_b (g_H \circ \zeta_{F^\dagger} \circ \mathrm{d}_{F^\dagger}(\ln m^r)_a) - \ln g_H \circ \zeta_{F^\dagger} \circ \mathrm{d}_{F^\dagger}(\ln m^r)_b$$

Therefore,

$$\widetilde{F}^a_b (g_H \circ \zeta_{F^\dagger} \circ \mathrm{d}_{F^\dagger}(\ln m^r)_a) = e^{C_{a\to b}} g_H \circ \zeta_{F^\dagger} \circ \mathrm{d}_{F^\dagger}(\ln m^r)_b$$

The collection of constant function $\mathbb{R} \hookrightarrow \widetilde{F}_a$ in each space F_a with $a \in \mathscr{A}$ forms a subobject of F, therefore $[\widetilde{F}](a) = \widetilde{F}_a/\mathbb{R}$, with associated quotient maps $[\widetilde{F}]^a_b$ is a presheave from $\mathscr{A}$ to finite vector spaces. A probability distribution $Q \in \widetilde{F}_a$ is in correspondance with $[Q] \in [\widetilde{F}](a)$, the section $b \in \widetilde{F}_a \to b/\sum b$ is a section which in turn provides the inverse map when passing to the quotient. The previous equation implies that,

$$[\widetilde{F}^a_b (g_H \circ \zeta_{F^\dagger} \circ \mathrm{d}_{F^\dagger}(\ln m^r)_a)] = [g_H \circ \zeta_{F^\dagger} \circ \mathrm{d}_{F^\dagger}(\ln m^r)_b]$$

Therefore, we have just shown that $\overline{\Delta\mathrm{BP}}([m^r]) = 0$ implies that

$$[g_H \circ \zeta_{F^\dagger} \circ \mathrm{d}_{F^\dagger}]([\ln m]^*) \in \lim[\widetilde{F}].$$

Define $b_a = (g_H \circ \zeta_{F^\dagger} \circ \mathrm{d}_{F^\dagger}([\ln m]^*))_a$. This statement can be reformulated as saying that

$$\left(\frac{b_a}{\sum_{x_a \in F_a} b_a}, a \in \mathscr{A}\right) \in \lim(\mathbb{P}F).$$

The previous equation implies that $\left(\frac{b_a}{\sum_{x_a \in F_a} b_a}, a \in \mathscr{A}\right)$ is a critical point of the Bethe Free Energy.

Conversely, if one starts with a critical point of the Bethe Free Energy, $p_a \in \mathbb{P}(F_a)$, $a \in \mathscr{A}$, then there exists a collection $(m^*_{a\to b};\, a,b \in \mathscr{A},\, b \leq a)$ such that

$$p = g_H \circ \zeta_{F^\dagger} \circ d_{F^\dagger}(m^*).$$

Since $p \in \lim \mathbb{P}F$, we have

$$\ln \widetilde{F}_b^a(p_a) - \ln p_b = 0$$

and

$$\mathrm{BP}(m^*) = m^*.$$

□

D.2 Proof of Proposition 3

We will follow the proof of Theorem 2.1 [21]. Let m^* be a fixed point of MP, i.e., $\Delta\mathrm{MP}(m^*) = 0$. This implies that

$$g_H \circ \zeta_{F^\dagger} \circ \mathrm{d}F^\dagger(\ln m^*) \in \lim \widetilde{F},$$

and that the associated normalized belief is in $\lim \mathbb{P}F$. By Lemma 2, therefore, m^* is a critical point of the Bethe Free Energy. □

E Link between belief propagation and message passage algorithms

When we study the functoriality of message passing algorithms, we consider the following operator:

$$\Delta\mathrm{MP}(l) = \delta_F \circ g_H \circ \zeta_{F^\dagger} \circ d_{F^\dagger}(l)$$

The messages l of MP are the $\ln m$ of BP, with for any $a,b \in \mathscr{A}$ such that $b \leq a$ and any $x_b \in F_b$, $m_{a\to b}(x_b) = e^{l_{a\to b}(x_b)}$. Let us now show that for any fixed point l^* of MP, $[m^*]$ is a fixed point of BP. We will then show that for any fixed point $[m]^*$ of BP, there exists $m^* \in [m]^*$ such that $\ln m^*$ is a fixed point of MP. Showing this justifies why the study of the critical points of the Belief Propagation algorithm is equivalent to the study of the MP algorithm considered in [21] and in this paper.

Let $\mathrm{MP}(l^*) = l^*$ then $\Delta\mathrm{MP}(l^*) = 0$ and so for any $b \leq a$, $F_b^a(g \circ \zeta_{F^\dagger} \circ d_{F^\dagger}(l)_a) = g \circ \zeta_{F^\dagger} \circ d_{F^\dagger}(l)_b$. Therefore, in particular $\Delta\mathrm{BP}(\ln m^*) = 0$ which implies that m^* is a fix point of BP.

For any message $\ln m_{a\to b} \in \widetilde{F}_b$, define $N_{a\to b}(\ln m_{a\to b}) = \ln m_{a\to b} - \ln \sum_{x_b} e^{\ln m_{a\to b}}$. Then, $N_{a\to b}(\ln m_{a\to b} + C) = N_{a\to b}(\ln m_{a\to b})$. Let us denote $N(\ln m) = (N_{a\to b}; a \to b)$.

Consider $[m]^*$ a fixed point of BP. Then, for any representative m^r such that $[m^r] = [m]^*$, we have $\Delta\mathrm{MP}(N(\ln m^r)) = 0$; therefore, $N(\ln m^r)$ is a fixed point of MP.

F Generalized Belief Propagation does not induce a functor with respect to natural transformations

The Generalized Belief Propagation update rule is as follows:

$$\ln m_{a\to b}^{t+1} = \ln m_{a\to b}^t + \ln \widetilde{F}_b^a\left(g_H \circ \zeta_{F^\dagger} \circ d_{F^\dagger}(\ln m^t)_a\right) - \ln\left(g_H \circ \zeta_{F^\dagger} \circ \mathrm{d}_{F^\dagger}(\ln m^t)_b\right)$$

Let us denote $\widetilde{\delta_F}$, as, for any $u \in \prod_{a\in\mathscr{A}} \mathbb{R}^{E_a}_{>0}$,

$$\widetilde{\delta_F}(u)_{a\to b} = \ln \widetilde{F}^a_b(u_a) - \ln u_b$$

Then,

$$\ln m^{t+1} = m^t + \widetilde{\delta_F} \circ g_H \circ \zeta_{F^\dagger} \circ d_{F^\dagger}(\ln m^t)$$

In particular, due to the nonlinearity of ln, $\widetilde{\delta}_F$ does not commute with ϕ if the preimage of at least one of the ϕ_a, for some $x \in G_a$, has cardinality at least 2.

G Proof of main theorems

G.1 Proof of Theorem 1

Now it is easy to show that $\phi_1 \circ \delta_F = \delta_G \circ \phi$ therefore by dualizing (as $d_F := (\delta_F)^* = d_{F^*}$) on ehas that $d_F \circ \phi_1^* = \phi^* \circ d_G$.

Let us now show that ζ also commutes to natural transformations for functors that have the same domain:

$$\zeta_{F^*} \circ \phi^*(v)(a) = \sum_{b\leq a} F^{a*}_b \phi^*(v)(b)$$

$$\sum_{b\leq a} F^{a*}_b \phi^*_b v_b = \sum_{b\leq a} \phi_b \circ F^{a*}_b v_b$$

and $\sum_{b\leq a} \phi_b \circ F^{a*}_b v_b = \sum_{b\leq a} G^a_b \circ {\phi_a}^* v_b = \sum_{b\leq a} \phi^*_a \circ G^{a*}_b v_b$ and ϕ_a is linear so, $\sum_{b\leq a} \phi^*_a \circ G^{a*}_b v_b = \phi^*_a \sum_{b\leq a} G^{a*}_b v_b$. So $\zeta_{F^*} \circ \phi^* = \phi^* \circ \zeta_{G^*}$.

$g_H : \bigoplus_{a\in\mathscr{A}} \widetilde{F}_a \to \bigoplus_{a\in\mathscr{A}} \widetilde{F}_a$ is defined by the following relationship:

$$d_x f_a = y \iff x = g_a(y) \tag{G.1}$$

where $f_a(q_a) = \sum q_a H_a - S(q_a) = \sum_\omega q_a(\omega)H_a(\omega) + \sum_\omega q_a(\omega)\ln q_a(\omega)$. The differential gives for the choice of scalar product

$$y_a = H_a + \ln q_a + 1$$

What we want to show is that there is $\widetilde{H} := (\widetilde{H}_a : G_a \to \mathbb{R}, a \in \mathscr{A})$ such that $\phi \circ g_H \circ \phi^\dagger = g_{\widetilde{H}}$. We must show it ponctually as we do not requiere any compatibility condition on the H_a: i.e. we need to show that, $\phi \circ g_{H_a} \circ \phi^\dagger = g_{\widetilde{H_a}}$. Let us first remark that for any $p_a \in \widetilde{F}_a$, $\phi^\dagger_a(p_a) = p_a \circ \phi_a$. Then,

$$\phi_a \circ g_{H_a} \circ \phi^\dagger_a(p)(y_a) = \sum_{x_a:\phi_a(x_a)=y_a} e^{-H_a(x_a)+p\circ\phi(x_a)-1} = e^{p(y_a)-1} \sum_{x_a:\phi_a(x_a)=y_a} e^{-H_a(x_a)}.$$

$$\phi_a \circ g_{H_a} \circ \phi^\dagger_a(p)(y_a) = \mathbf{1}[y_a \in \operatorname{im}\phi_a] e^{p(y_a)-1-\widetilde{H}_a}.$$

By adopting the convention that $e^{-\infty} = 0$, we write:

$$\phi_a \circ g_{H_a} \circ \phi^\dagger_a(p)(y_a) = e^{p(y_a)-1-\widetilde{H}_a} = g_{\widetilde{H}_a}.$$

W proved that if there is a natural transformation of measurable map between $\phi : F \to G$ then,

$$\Delta MP_G = \phi \circ \Delta MP_F \circ \phi^\dagger$$

and so one deduces from the fact that $g^{\langle \cdot,\cdot \rangle}_{\widetilde{H}_a} = \phi_a \circ g^{\langle \cdot,\cdot \rangle}_{\widetilde{H}_a} \circ \phi^\dagger_{a,\langle \cdot,\cdot \rangle}$ that

$$\Delta MP^{\langle \cdot,\cdot \rangle}_G = \phi \circ \Delta MP^{\langle \cdot,\cdot \rangle}_F \circ \phi^\dagger$$

.

G.2 Proof of Theorem 2

Consider $\phi : F \to G$ a natural transformation, then $\Delta\mathrm{MP}(\phi) = (\phi, \psi)$ and by Theorem1,

$$\Delta\mathrm{MP}(G)(l, \psi(H)) = \phi \circ \Delta\mathrm{MP}_H(\phi^\dagger(l))$$

Furthermore, for any $a \in \mathscr{A}$ and $y_a \in G_a$, the mapping $H \mapsto \psi(H)_a(y_a)$ is convex and therefore continuous. Since $\mathscr{A}$ is finite and each G_a is finite, ψ is a continuous map. We just showed that $\Delta\mathrm{MP}(\phi) : \Delta\mathrm{MP}(F) \to \Delta\mathrm{MP}(G)$ is a morphism of **MPA**.

Let us now show that ΔMP is functorial. Let F, G, L be three presheaves of $\hat{\mathscr{A}}_f$, and let $\phi : F \to G$, $\phi_1 : G \to L$ be two natural transformations. Then,

$$\forall a \in \mathscr{A}, z_a \in L_a, \quad \Delta\mathrm{MP}(\phi_1) \circ \Delta\mathrm{MP}(\phi)(z_a) = -\ln \sum_{y_a \in G_a : \phi_{1,a}(y_a) = z_a} e^{-\psi_a(y_a)}$$

$$-\ln \sum_{\substack{y_a \in G_a \\ \phi_{1,a}(y_a) = z_a}} e^{-\psi_a(y_a)} = -\ln \sum_{\substack{y_a \in G_a \\ \phi_{1,a}(y_a) = z_a}} \sum_{\substack{x_a \in F_a \\ \phi_a(x_a) = y_a}} e^{-H_a(x_a)}$$

$$= -\ln \sum_{\substack{x_a \in F_a \\ \phi_1 \circ \phi_a(x_a) = z_a}} e^{-H_a(x_a)}$$

Therefore ΔMP is a functor.

□

G.3 Proof of Theorem 3

Let F, G be two presheaves of $\hat{A}_f$, let $\phi : F \to G$ be a epimorphism, i.e. a natural transformation where each map is surjective. Let $\mathrm{MP}(\phi) = (\phi, \psi)$ with,

$$\forall a \in \mathscr{A}, y_a \in G_a, \quad \psi(H)_a(y_a) = -\ln \sum_{x_a \in F_a : \phi_a(x_a) = y_a} e^{-H_a(x_a)}.$$

By Theorem 2, $\phi \circ \Delta\mathrm{MP}_{F,H} \circ \phi^\dagger = \Delta\mathrm{MP}_{G,\psi(H)}$.

$$\phi \circ \mathrm{MP}_{F,H} \circ \phi^\dagger(l) = \phi \circ \phi^\dagger(l) + \phi \circ \mathrm{MP}_{F,H} \circ \phi^\dagger(l)$$

For the choice of scalar product $\langle \cdot,\cdot \rangle^\phi$, by Lemma 1, $\phi \circ \phi^\dagger = \mathrm{id}$, therefore,

$$\phi \circ \mathrm{MP}_{F,H} \circ \phi^\dagger(l) = l + \Delta\mathrm{MP}_{F,\psi(H)}(l) = \mathrm{MP}_{G,\psi(H)}(l),$$

and therefore $\mathrm{MP}(\phi) : \mathrm{MP}(F) \to \mathrm{MP}(G)$ is a morphism of MPA; functoriality is a consequence of Theorem 2, as $\Delta\mathrm{MP}(\phi) = \mathrm{MP}(\phi)$.