



HAL
open science

Inference on diagrams in the category of Markov kernels (Extended abstract)

Grégoire Sergeant-Perthuis, Nils Ruet

► **To cite this version:**

Grégoire Sergeant-Perthuis, Nils Ruet. Inference on diagrams in the category of Markov kernels (Extended abstract). 2024. hal-04527780

HAL Id: hal-04527780

<https://hal.sorbonne-universite.fr/hal-04527780>

Preprint submitted on 30 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

Inference on diagrams in the category of Markov kernels (Extended abstract)

Grégoire Sergeant-Perthuis*

Nils Ruet †

Graphical models are widely used families of probability distributions that capture conditional independence relations between a collection of variables $X_i, i \in I$; celebrated examples are Hidden Markov models and Bayesian Networks [1]. Graphical models are built from directed and undirected graphs $G = (I, A)$ where nodes $i \in I$ are uniquely identified with the variables X_i . Inference in graphical models ultimately boils down to inference for an undirected graphical model, achieved through the Belief Propagation algorithm [2]. Such Inference constitutes a specific instance of variational inference as it revolves around a free energy termed the Bethe free energy [2]. Adopting a variational inference perspective for graphical models has facilitated the extension of the Belief Propagation algorithm to encompass broader classes of probability distributions, enabling the accommodation of interactions among more than 2 variables in contrast to traditional graphical models (factor graphs [3]); this is achieved through the introduction of the Kikuchi free energies [4]. Let us denote $\mathbf{Mes}^f, \mathbf{Kern}^f$, the categories with objects finite measurable spaces and respectively with morphisms measurable maps and the second Markov Kernels (stochastic matrices). \mathbf{Mes}^f can be seen as a subcategory of \mathbf{Kern}^f . As exhibited in [5–7], what underlies variational inference for those classes of probability distributions are presheaves from a finite poset to \mathbf{Mes}^f , which morphisms are epimorphisms. We will call them the ‘graphical’ presheaves. Our contribution is to extend the Generalized Belief Propagation [5] to any presheaf from a finite poset to \mathbf{Kern}^f . This work is contained in Chapter 9 of [8] and Appendix 1 of unpublished [9], where we consider the more general problem of optimizing a collection of cost functions over a presheaf of signals.

1. Motivation and related work

Consider a collection of agents represented by vertices $i \in I$ that can communicate their beliefs to neighboring vertices ∂i through undirected edges $e \in A$. Each agent has its own representation of its environment, denoted by E_i . They can share their beliefs with neighboring nodes $j \in \partial i$ through a measurable map $f_e^i : E_i \rightarrow E_e$. Graphical models and their extensions do not allow us to account for such heterogeneity in the way each agent models their environment. Such setting is better captured by cellular sheaves [10] and applications [11], important examples of which are Sheaf Neural Networks [12], are limited to functors from the poset associated to a graph ($i \leq e \iff i \in e$) to the category of finite vector spaces \mathbf{Vect}^f . We are interested in the more general case where beliefs transfer through a hierarchy, i.e. a poset, and we provide an algorithm for inference in such case where Sheaf Neural Networks can’t be used; by convention, we consider presheaves instead of functors: ‘orders’ are given top-down. More generally, cellular sheaves are restricted to the face poset of a cell complex and hence don’t apply to all hierarchies and therefore not to our case.

2. Free energy for poset shaped diagrams in \mathbf{Kern}^f and message passing algorithm

Definition 1 (Graphical presheaves). Let I be a finite set and $\mathcal{A} \subseteq \mathcal{P}(I)$ be a sub-poset of the powerset of I . Let $E_i, i \in I$ are finite sets. For $a \in \mathcal{A}$ $E_a := \prod_{i \in a} E_i$, let $F(a) := E_a$, and for $b \subseteq a$, let $F_b^a : E_a \rightarrow E_b$

*LCQB, Sorbone universit , Paris, France, gregoire.sergeant-perthuis@sorbonne-universite.fr

†CIAMS, Universit  Paris-Saclay, Orsay, France

be the projection map from $\prod_{i \in a} E_i$ to $\prod_{i \in b} E_i$. F is called a graphical presheaf from \mathcal{A} to \mathbf{Mes}^f .

For \mathcal{A} a finite poset, the ‘zeta-operator’ of \mathcal{A} , denoted ζ , from $\bigoplus_{a \in \mathcal{A}} \mathbb{R}$ to $\bigoplus_{a \in \mathcal{A}} \mathbb{R}$ is defined as, for any $\lambda \in \bigoplus_{a \in \mathcal{A}} \mathbb{R}$ and any $a \in \mathcal{A}$, $\zeta(\lambda)(a) = \sum_{b \leq a} \lambda_b$. ζ is invertible [13], we denote μ its inverse; its matrix expression $(\mu(a, b), b \leq a)$ defines the Möbius function of \mathcal{A} . Let F be a presheaf from \mathcal{A} to \mathbf{Kern}^f ; $F_b^a : F(a) \rightarrow F(b)$ is denoted element-wise as $F_b^a(\omega_b | \omega_a)$, with $\omega_b \in F(b), \omega_a \in F(a)$. It induces a presheaf \tilde{F} from \mathcal{A} to \mathbf{Vect}^f , where $\tilde{F}_b^a : \mathbb{P}(F(a)) \rightarrow \mathbb{P}(F(b))$ is the linear map that sends probability distributions $p \in \mathbb{P}(F(a))$ to $F_b^a \circ p$. Following [5], we introduce a free energy $\mathcal{F}(Q) = \sum_{a \in \mathcal{A}} c(a) (\mathbb{E}_{Q_a}[H_a] - S(Q_a))$; $c(a) = \sum_{b \geq a} \mu(b, a)$ is the generalization of the inclusion-exclusion formula associated to \mathcal{A} . $S(Q_a) = -\sum_{\omega_a \in F(a)} Q_a(\omega_a) \ln Q_a(\omega_a)$ is the entropy of Q_a . We propose to solve $\inf_{Q \in \lim \tilde{F}} \mathcal{F}(Q)$. \tilde{F}^* is the functor obtained by dualizing the morphisms \tilde{F}_b^a , i.e. $\tilde{F}_b^a : \tilde{F}(b)^* \rightarrow \tilde{F}(a)^*$ sends linear maps $l_b : \tilde{F}(b) \rightarrow \mathbb{R}$ to $l_b \circ \tilde{F}_b^a : \tilde{F}(a) \rightarrow \mathbb{R}$.

For a functor G from \mathcal{A} to \mathbb{R} -vector spaces, we define μ_G as, for any $a \in \mathcal{A}$ and $v \in \bigoplus_{a \in \mathcal{A}} G(a)$, $\mu_G(v)(a) = \sum_{b \leq a} \mu(a, b) G_b^a(v_b)$. Let us define the function $FE : \prod_{a \in \mathcal{A}} \mathbb{P}(E_a) \rightarrow \prod_{a \in \mathcal{A}} \mathbb{R}$ as $FE(Q) = (\mathbb{E}_{Q_a}[H_a] - S_a(Q_a), a \in \mathcal{A})$, which sends a collection of probability measures over \mathcal{A} to their Gibbs free energies. For any $Q \in \prod_{a \in \mathcal{A}} \mathbb{P}(E_a)$, let us denote $d_Q FE$ as the differential of FE at the point Q .

Theorem 1. *Let \mathcal{A} be a finite poset, let F be a presheaf from \mathcal{A} to \mathbf{Kern}^f . Let $H_a : F(a) \rightarrow \mathbb{R}$ be a collection of (measurable) functions. The critical points of \mathcal{F} are the $Q \in \lim \tilde{F}$ such that,*

$$\mu_{\tilde{F}^*} d_Q FE|_{\lim \tilde{F}} = 0 \quad (1)$$

The message-passing algorithm we consider is Algorithm 1; it specializes to the General Belief Propagation for graphical presheaves. For two elements $a, b \in \mathcal{A}$, such that $b \leq a$, two types of messages are considered: top-down messages $m_{a \rightarrow b} \in \mathbb{R}^{F(b)}$ and bottom-up messages $n_{b \rightarrow a} \in \mathbb{R}^{F(a)}$.

Algorithm 1: Message passage algorithm for presheaves from \mathcal{A} to \mathbf{Kern}^f

Data: Initialization: $(m_{a \rightarrow b}^0 \in \mathbb{R}^{F(b)}, b, a \in \mathcal{A} \text{ s.t. } b \leq a)$, a poset \mathcal{A} , a presheaf $F : \mathcal{A} \rightarrow \mathbf{Kern}^f$;

- 1 **for** $t \leq T$ **do**
- 2 **for** $a \in \mathcal{A}, b \in \mathcal{A}$ such that $b \leq a$ **do**
- 3 $\forall \omega_a \in F(a), n_{b \rightarrow a}(\omega_a) \leftarrow \prod_{\substack{c: b \leq c \\ c \not\leq a}} \sum_{\omega'_b \in F(b)} m_{c \rightarrow b}(\omega'_b) \cdot F_b^a(\omega'_b | \omega_a)$
- 4 **end**
- 5 **for** $a \in \mathcal{A}, b \in \mathcal{A}$ such that $b \leq a$ **do**
- 6 $b_a = e^{-H_a} \prod_{\substack{b \in \mathcal{A} \\ b \leq a}} n_{b \rightarrow a}$
- 7 $p_a = \frac{b_a}{\sum_{\omega_a} b_a(\omega_a)}$
- 8 $m_{a \rightarrow b} \leftarrow m_{a \rightarrow b} \cdot \frac{\tilde{F}_b^a(p_a)}{p_b}$
- 9 **end**
- 10 **end**

A criterion to stop the algorithm is when the beliefs do not change, i.e., when $p_a^{t+1} \approx p_a^t$. The fixed points of the previous message-passing algorithm correspond to critical points of \mathcal{F} over $\lim F$ (Corollary of Theorem 2.2 [9]v2). Theorem 1 differs from a similar characterization of critical points of a free energy for specifications in [14] by the fact that the $\mu_{\tilde{F}^*}$ and $\lim \tilde{F}$ are applied to the same presheaf \tilde{F} and not two different presheaves/functors (G, F) .

References

- [1] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, 1988.
- [2] J. S. Yedidia, W. T. Freeman, and Y. Weiss, *Understanding belief propagation and its generalizations*, p. 239–269. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003.
- [3] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [4] J. S. Yedidia, W. T. Freeman, and Y. Weiss, “Generalized belief propagation,” in *Proceedings of the 13th International Conference on Neural Information Processing Systems, NIPS’00*, (Cambridge, MA, USA), p. 668–674, MIT Press, 2000.
- [5] J. Yedidia, W. Freeman, and Y. Weiss, “Constructing Free-Energy Approximations and Generalized Belief Propagation Algorithms,” *IEEE Transactions on Information Theory*, vol. 51, pp. 2282–2312, July 2005.
- [6] O. Peltre, “Message passing algorithms and homology,” 2020. Ph.D. thesis, Link.
- [7] O. Peltre, “A homological approach to belief propagation and Bethe approximations,” in *International Conference on Geometric Science of Information*, pp. 218–227, Springer, 2019.
- [8] G. Sergeant-Perthuis, *Intersection property, interaction decomposition, regionalized optimization and applications*. PhD thesis, Université de Paris, 2021. Link.
- [9] G. Sergeant-Perthuis, “Regionalized optimization,” 2022. <https://arxiv.org/abs/2201.11876>.
- [10] J. Curry, *Sheaves, cosheaves and applications*. PhD thesis, The University of Pennsylvania, 2013. arXiv:1303.3255.
- [11] J. Hansen and R. Ghrist, “Distributed optimization with sheaf homological constraints,” in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 565–571, 2019.
- [12] C. Bodnar, F. D. Giovanni, B. P. Chamberlain, P. Lio, and M. M. Bronstein, “Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in GNNs,” in *Advances in Neural Information Processing Systems* (A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, eds.), 2022.
- [13] G.-C. Rota, “On the foundations of combinatorial theory I. Theory of Möbius functions,” *Probability theory and related fields*, vol. 2, no. 4, pp. 340–368, 1964.
- [14] G. Sergeant-Perthuis, “Compositional statistical mechanics, entropy and variational inference,” 2024. <https://arxiv.org/abs/2403.16104>.