Homological Algebra for Message-Passing Algorithms

Olivier Peltre

Paris University – IMJ-PRG
February 2020

Introduction

The problem of describing the statistics of a large number $x_i, x_j, \dots$ of interacting random variables emerged in physics with Boltzmann's efforts to lay principles of thermodynamics on statistical grounds. High dimensional statistics are now also expected to provide with reasonable and tractable models in artificial intelligence and biology. Aimed at modelling the emergence of collective behaviours in large assemblies of constituents, the prism of statistics hence shows deep analogies between atoms in a crystal, and neurons in a network.

A probability distribution $p(x)$ on the joint variable $x = (x_j)_{j\in\Omega}$ is usually assumed to capture all collective phenomena, although a dimensional curse prohibits the computation of expectation values. Their local effects on a small subset $\alpha \subseteq \Omega$ of variables may nonetheless be estimated, as statistics of the local joint variable $x_\alpha = (x_i)_{i\in\alpha}$ only involve its marginal distribution $p_\alpha(x_\alpha)$, which lives in a space of reasonable dimension.

Accessing marginals is a crucial step of statistical learning. Usually appearing in the gradient of a loss function, they are necessary to guide the update of model parameters. *Message-passing algorithms* estimate marginals through a parallelised and asynchronous computing scheme, in which a collection of local units communicate until they eventually reach a consensual state. Such algorithms best leverage on the local structure of the probabilistic model, described below.

Gibbs random fields[1] are probabilistic models associated to a collection X of subsets $\alpha, \beta, \gamma, \dots$ of Ω over which the global distribution $p(x)$ factorises as a product of local functions. We write $p \in \mathcal{G}(X)$ when there exists a collection of positive factors (f_α) such that:

$$p(x) = \frac{1}{Z} \prod_{\alpha\in X} f_\alpha(x_\alpha) \tag{1}$$

Distributions of this form are more often called *graphical models* in the computer science literature. The hypergraph $X \subseteq \mathcal{P}(\Omega)$ is then represented by the so-called factor graph, depicted in figure 1, formed by joining variable nodes (x_i) with their associated factor nodes (f_α). Positivity of the factors implies that p may be written as the *exponential family*:

$$p = \frac{\mathrm{e}^{-H}}{Z} \qquad \text{where} \qquad H(x) = \sum_{\alpha\in X} -\ln f_\alpha(x_\alpha) \tag{2}$$

In physics, p is called the Gibbs state associated to the *hamiltonian* H, also called *total energy*. One should also be prepared to encounter the name of *energy-based model* and its acronym EBM, in the context of Boltzmann machines for instance.

[1] Or *Gibbs distributions*, which are also *Markov random fields* according to the Hammersley-Clifford correspondence. Factorisability however yields a finer characterisation of $\mathcal{G}(X)$ than Markov properties, hence the preferred terminology.

(a) (b)

Figure 1: (a) factor graph and (b) Venn representations of $X \subseteq \mathcal{P}(\Omega)$.

Our main contribution is to view Gibbs random fields as *homology classes* of factors. The 1-chains $(m_{\alpha\beta})$ resolving the ambiguity of the parametrisation (1) of p by the factors (f_α) are called *messages*. They may be iterated upon by following the sum-product update rule of belief propagation for instance. The differential structure is however best presented additively, at the level of energies or log-likelihoods.

The following is an abstract of a PhD work done under the kind supervision of D. Bennequin, to which I am greatly indebted. I also wish to thank G. Sergeant-Perthuis and J.P. Vigneaux, for the early and ongoing fruitful collaborations without which this work would not have been possible.

The initial data we shall start with is a collection $(E_i)_{i\in\Omega}$ of finite sets, indexed by a finite set Ω, describing the configuration space of each individual variable x_i for $i \in \Omega$. Binary variables $E_i = \{\pm 1\}$ already cover many applications, from the Ising model to Hopfield networks [6].

1 Statistical Systems and Functors

Regions. The powerset $\big(\mathcal{P}(\Omega), \supseteq\big)$ is a category for its partial order structure. We write:

$$\alpha \to \beta \quad \Leftrightarrow \quad \alpha \supseteq \beta$$

A collection of subsets[2] $X \subseteq \mathcal{P}(\Omega)$ has the same arrows as a subcategory. Such a *regionalisation* X is assumed chosen pragmatically. Note that X could for instance be a lattice (Ising), a fully-connected graph (Hopfield) or more generally, a covering of Ω by subsets of reasonable size.

Microstates. Identify E with the functor $E : \mathcal{P}(\Omega) \to \mathbf{Set}$ lifted to subsets by:

$$E_\alpha = \prod_{i\in\alpha} E_i$$

For every $\beta \subseteq \alpha$, denote by $\pi^{\beta\alpha} : E_\alpha \to E_\beta$ the canonical projection, which forgets the microstate of variables outside β. The collection (E, π) is also called a *projective system* of sets. Viewing Ω as a discrete topological space, the functor $E : \mathcal{P}(\Omega) \to \mathbf{Set}$ moreover forms a *sheaf* of sets over Ω.

Observables. Denote by $A = \mathbb{R}^{\mathbb{E}}$ the functor $A : \mathcal{P}(\Omega)^{op} \to \mathbf{Alg}$ of functions over microstates:

$$A_\alpha = \mathbb{R}^{E_\alpha} = \bigotimes_{i\in\alpha} A_i$$

For every $\beta \subseteq \alpha$, denote by $j_{\alpha\beta} : A_\beta \to A_\alpha$ the pull-back of $\pi^{\beta\alpha}$, which canonically extends a function on E_β to E_α. Note that we shall keep extensions implicit and simply write $f_\beta(x_\beta) = f_\beta(\pi^{\beta\alpha}(x_\alpha))$. The pair (A, j) is also called an *inductive system* of algebras.

[2] A collection of subsets is often called a *hypergraph*, as combinatorial graphs are a particular case where subsets have cardinalities 1 and 2. The term *regionalisation* comes from Yedidia et.al. [20]

Measures. Denote by $A^* : \mathcal{P}(\Omega) \to \mathbf{Vect}$ the functor of linear forms over observables.

$$A^*_\alpha = L(A_\alpha, \mathbb{R}) \simeq \mathbb{R}^{E_\alpha}$$

For every $\beta \subseteq \alpha$, denote by $\Sigma^{\beta\alpha} : A^*_\alpha \to A^*_\beta$ the adjoint of $j_{\alpha\beta}$, push-forward of measures by $\pi^{\beta\alpha}$. The map $\Sigma^{\beta\alpha}$ performing partial integration over the state of variables in $\alpha \smallsetminus \beta$, is often called marginal projection. The pair (A, Σ) forms a *projective system* of vector spaces.

Statistical States. Denote by $\bar\Delta : \mathcal{P}(\Omega) \to \mathbf{Conv}$ the restriction of A^* to probability densities:

$$\bar\Delta_\alpha = \left\{ p_\alpha \in A^*_\alpha \;\middle|\; \langle p_\alpha \,|\, 1 \rangle = 1 \quad \text{and} \quad \forall f_\alpha \in A_\alpha \;\; \langle p_\alpha \,|\, f_\alpha^2 \rangle \geq 0 \right\}$$

We shall also denote by Δ_α the interior of $\bar\Delta_\alpha$ formed by non-degenerate probability densities. The pair (Δ, Σ) forms a *projective system* of convex topological spaces.

2 Energy Conservation and Homology

Observable Fields. There is a canonical lifting of $A : X^{op} \to \mathbf{Alg}$ to a functor on the nerve[3] of X which we still denote by $A : N(X) \to \mathbf{Alg}$, defined by letting for all $\bar\alpha = \alpha_0 \supseteq \ldots \supseteq \alpha_p \in N_p(X)$:

$$A_{\alpha_0 \ldots \alpha_p} \simeq A_{\alpha_p}$$

Note that whenever $\bar\alpha \in N_k(X)$ is a subface of $\bar\beta \in N_p(X)$ with $k \leq n$, one has $\alpha_k \supseteq \beta_p$ so that $A_{\bar\beta}$ canonically injects into $A_{\bar\alpha}$. The graded vector space $A_\bullet(X)$ is then defined by letting for all $p \in \mathbb{N}$:

$$A_p(X) = \bigoplus_{\bar\alpha \in N_p(X)} A_{\bar\alpha}$$

A 0-field $(u_\alpha) \in A_0(X)$ is hence a choice of an observable $u_\alpha \in A_\alpha$ for every $\alpha \in X$.

Boundary Operator. The simplicial structure of $N(X)$ yields a chain complex $(A_\bullet(X), \delta)$. This construction was considered by Grothendieck and Verdier in SGA-4-V [18] under the name of *canonical projective resolution for presheaves.* The action of the first degree boundary $\delta : A_1(X) \to A_0(X)$ on a flux $\big(\varphi_{\alpha\beta}(x_\beta)\big)$ is given by:

$$(\delta\varphi)_\beta(x_\beta) = \sum_{\alpha' \supseteq \beta} \varphi_{\alpha'\beta}(x_\beta) - \sum_{\beta \supseteq \gamma'} \varphi_{\beta\gamma'}(x_{\gamma'})$$

Energy Conservation. Given two collections of *interaction potentials* (u_α) and (u'_α) in $A_0(X)$:

$$\exists\, \varphi \quad \text{s.t.} \quad u' = u + \delta\varphi \quad \Rightarrow \quad \sum_\alpha u_\alpha = \sum_\alpha u'_\alpha$$

The converse implication, proved in [12, 13], strongly relies on the interaction decomposition theorem [9, 7] which characterises direct sum splittings of each A_α as $\bigoplus_{\beta \subseteq \alpha} Z_\beta$. It expresses that the homology class of (u_α) is completely characterised by its *total energy*:

$$H_\Omega(x_\Omega) = \sum_\alpha u_\alpha(x_\alpha)$$

[3]The p-nerve $N_p(X)$ of X is the set of p-chains $\bar\alpha = \alpha_0 \supseteq \ldots \supseteq \alpha_p$ in X. It should not be confused with the *Čech nerve* of X viewed as a covering, considered in a similar setting in [1], whose construction is related to ours in [2]. Given any category $\mathbf{C}$, its nerve $N_\bullet(\mathbf{C})$ is a simplicial set, occuring for instance in the construction of classifying spaces [16].

3 Marginal Consistency and Cohomology

Measure Fields. The functor of measures $A^* : X \to \mathbf{Vect}$ reciprocally extends to the nerve of X as a functor $A^* : N(X)^{op} \to \mathbf{Vect}$ by duality. We shall also denote by:

$$\Delta_0(X) = \prod_{\alpha \in X} \Delta_\alpha$$

the convex subset of $A_0^*(X)$ formed by non-degenerate probability densities.

Differential. Measure fields naturally form a cochain complex $(A_\bullet^*(X), d)$ with $d = \delta^*$. The action of the differential $d : A_0^*(X) \to A_1^*(X)$ on a field (q_α) is given by:

$$(dq)_{\alpha\beta}(x_\beta) = q_\beta(x_\beta) - \sum_{y \in E_{\alpha \smallsetminus \beta}} q_\alpha(x_\beta, y)$$

Marginal Consistency. A collection of measures $(q_\alpha) \in A_0^*(X)$ is said *consistent* when q_β is the marginal of q_α for every $\beta \subseteq \alpha$. Hence (q_α) is consistent when it is a 0-cocycle, satisfying $dq = 0$, i.e. $q_\beta = \Sigma^{\beta\alpha}(q_\alpha)$ for all $\beta \subseteq \alpha$. We denote by:

$$\Gamma(X) = \Delta_0(X) \cap \mathrm{Ker}(d)$$

the space of consistent positive probability densities, which is the *projective limit* of the system (Δ, Σ).

4 Message-Passing: Dynamics and Equilibria

Local Gibbs States. The Legendre duality[4] between energy and statistical state translates to local fields, associating to interaction potentials $(u_\alpha) \in A_0(X)$ the collection of *beliefs* $(q_\alpha) \in \Delta_0(X)$ by:

$$q_\alpha = \frac{\mathrm{e}^{-U_\alpha}}{Z_\alpha} \qquad \text{where} \qquad U_\alpha = \sum_{\beta \subseteq \alpha} u_\beta \tag{3}$$

We shall write $U = \zeta \cdot u$ to denote the bijective correspondence [15] between local hamiltonians and interaction potentials. We also write $q = [\mathrm{e}^{-U}]$, letting the bracket denote normalisation.

Statistical Diffusion. We introduce the ordinary differential equation in $A_0(X)$ defined by:

$$\frac{du}{dt} = \delta\big(\mathcal{D}(\zeta \cdot u)\big) \qquad \text{with} \qquad \mathcal{D}(U)_{\alpha\beta} = U_\beta + \ln\big(\Sigma^{\beta\alpha}(\mathrm{e}^{-U_\alpha})\big) \tag{4}$$

We shall rewrite (4) as $\dot{u} = \delta\Phi(u)$ where $\Phi : A_0(X) \to A_1(X)$ is the smooth *flux functional*[5] $\Phi = \mathcal{D} \circ \zeta$. We show in [13] that the *belief propagation* algorithm introduced by Gallager [5], popularised by Pearl [11] and later generalised in [20], is equivalent to the time-step-1 explicit Euler scheme of (4):

$$\mathrm{BP} : \quad u(t+1) \leftarrow u(t) + \delta\Phi(u(t)) \tag{5}$$

The witnessed unstable behaviour of BP, sometimes compared to a kind of «phase transition» [14], should be expected from the Lipschitz bound of Φ tending to 1 as beliefs go the boundary of $\Delta_0(X)$. We hence strongly advise for the use of smaller-time-step integrators BP_λ of the ODE (4).

[4] The free energy functional $F(H) = -\ln \Sigma\, \mathrm{e}^{-H}$ is conjugated to the Shannon entropy functional $S(p) = -\Sigma\, p \ln(p)$. This Legendre duality puts in correspondence hamiltonians with statistical states at equilibrium.

[5] Extending the zeta transform ζ and its inverse $\mu = \zeta^{-1}$ to the higher degrees of $A_\bullet(X)$, we propose in [13] another flux functional $\phi = \mu \circ \mathcal{D} \circ \zeta$, correcting the redundancies counted in the generalised BP algorithm of [20].

Equilibria. Stationary states of (4) lie at the intersection of a homology class $[h] = h + \delta A_1(X)$ with a manifold $\mathcal{Z}(X)$ of consistent interaction potentials, mapped onto[6] $\Gamma(X)$ by the smooth application $A_0(X) \to \Delta_0(X)$ defining local Gibbs states (3):

$$\mathcal{Z}(X) = \big\{ u \in A_0(X) \mid \mathcal{D}(\zeta \cdot u) = 0 \big\}$$

To approximate the marginals (p_α) of a global probability distribution $p_\Omega = [\mathrm{e}^{-H_\Omega}]$, message-passing algorithms hence seek for consistent beliefs $(q_\alpha) \in \Gamma(X)$ deriving from interaction potentials $(u_\alpha) \in [h]$ that preserve the total energy H_Ω.

Bifurcations. In general[7], multiple stationary states coexist in a given homology class, as numerical experiments have already shown [19, 8]. We relate this phenomenon to *singularities* of the homological projection $\mathcal{Z}(X) \to \mathrm{H}(X)$, which occur when $T_u\mathcal{Z}(X) \cap \delta A_1(X)$ is non-trivial.

Denoting by $\mathcal{B}(X) \subseteq TA_0(X)$ the sub-bundle of tangent vectors spanned by $\delta A_1(X)$ and restricted above $\mathcal{Z}(X)$, consider the linearised diffusion flow[8]:

$$L = \delta \nabla \zeta \;\in\; \mathrm{Hom}\big(\mathcal{B}(X), \mathcal{B}(X)\big)$$

Following Thom [17], the corank of L induces a decomposition of $\mathcal{Z}(X)$ as the stratified space $\bigsqcup_k \mathcal{S}_k$, where L is invertible on the dense subspace $\mathcal{S}_0$, complement of the analytic subspace $\bar{\mathcal{S}}_1$ of singularities. Explicit examples of such bifurcations can be exhibited on simple graphs with more than two loops, e.g. on the join of two triangles by a vertex or an edge [13].

Conclusion

The algebraic and geometric structure of Gibbs random fields allows for a better understanding of the purpose and behaviour of message passing algorithms. Although non-linear, the present theory shows many similarities with the usual theory of harmonic forms. Its main difference is the central role played by the zeta transform in defining local hamiltonians from interaction potentials.

We would also like to mention the equivalence of message-passing stationary states with critical points of the *Bethe free energy* functional [4, 10], a local approximation of the global free energy. In this dual picture, flux terms $\delta\varphi$ appear as Lagrange multipliers for the consistency constraint $dq = 0$ imposed on beliefs [20, 13]. These two point of views shed a different light on the benefits brought by combinatorics to statistics.

We believe that viewing the transport equation (4) as an ODE is extremely important for stability concerns. As of today, the few simulations we have run showed dramatic change of behaviour when turning the time steps from 1 to 0.1. More experiments are called for and should soon be provided. Note that shorter time steps appear as exponents in the usual multiplicative formalism, this perhaps explains why such a simple improvement of belief propagation came unnoticed.

An extension of the present constructions to quantum statistics may also be carried out naturally. In the quantum setting, instead of microstates, one would start with a functor of non-commutative observables algebras $A : X^{op} \to \mathbf{Alg}$. The non-linear correspondence (3) can be extended to the case of quantum states represented by non-degenerate density matrices. One may then estimate the local effects of the time evolution operator e^{itH} and the marginals of the density matrix $\mathrm{e}^{-\theta H}$ by message-passing algorithms eitherwise.

[6] The action of region-wise additive constants $\mathbb{R}_0(X)$ on $\mathcal{Z}(X)$ spans the reciprocal image of $\Gamma(X)$ under (3).

[7] When X is an acyclic graph, BP is known to converge in finite time to the exact marginals of the global Gibbs state. This result may be extended to a wider class of *acyclic* hypergraphs, but fails when X is a graph with loops.

[8] The differential $\nabla = \mathcal{D}_*$ acts on local hamiltonians by $\nabla(U)_\beta = U_\beta - \mathbb{E}_{q_\alpha}[U_\alpha \mid \beta]$. It is the adjoint of δ for a metric induced by the consistent local Gibbs states (q_α).

References

[1] S. Abramsky and A. Brandenburger, *The Sheaf-theoretic structure of non-locality and contextuality*, New Journal of Physics, 13 (2011).

[2] D. Bennequin, O. Peltre, and G. Sergeant-Perthuis, *Extrafine Sheaves.* preprint, 2019.

[3] D. Bennequin, O. Peltre, G. Sergeant-Perthuis, and J. Vigneaux, *Informations, Energies and Messages.* preprint, 2019.

[4] H. A. Bethe and W. L. Bragg, *Statistical Theory of Superlattices*, Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences, 150 (1935), pp. 552–575.

[5] R. G. Gallager, *Low-Density Parity-Check Codes*, MIT Press, 1963.

[6] J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci. USA, 79 (1982), pp. 2554–2558.

[7] H. G. Kellerer, *Maßtheoretische Marginalprobleme*, Mathematische Annalen, 153 (1964), pp. 168–198.

[8] C. Knoll and F. Pernkopf, *On Loopy Belief Propagation – Local Stability Analysis for Non-Vanishing Fields*, in Uncertainty in Artificial Intelligence, 2017.

[9] F. Matúš, *Discrete Marginal Problem for Complex Measures*, Kybernetika, 24 (1988), pp. 36–46.

[10] T. Morita, *Cluster Variation Method of Cooperative Phenomena and its Generalization I*, Journal of the Physical Society of Japan, 12 (1957), pp. 753–755.

[11] J. Pearl, *Reverend Bayes on Inference Engines: A Distributed Hierachical Approach*, in AAAI-82 Proceedings, 1982.

[12] O. Peltre, *A Homological Approach to Belief Propagation and Bethe Approximations*, in Geometric Science of Information, 4th International Conference GSI 2019, Springer, 2019.

[13] ——, *Homology of Message-Passing Algorithms.* PhD preprint, 2020.

[14] W. Ping and A. Ihler, *Belief Propagation in Conditional RBMs for Structured Prediction*, in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, vol. 54, 2017, pp. 1141–1149.

[15] G.-C. Rota, *On the Foundations of Combinatorial Theory - I. Theory of Möbius Functions*, Z. Warscheinlichkeitstheorie, 2 (1964), pp. 340–368.

[16] G. Segal, *Classifying spaces and spectral sequences*, Publications Mathématiques de l'IHÉS, 34 (1968), pp. 105–112.

[17] R. Thom, *Les Singularités des applications différentiables*, Annales de l'institut Fourier, 6 (1956), pp. 43–87.

[18] J.-L. Verdier and A. Grothendieck, *V: Cohomologie dans les Topos*, SGA-4, 2 (1972).

[19] Y. Weiss, *Belief propagation and revision in networks with loops*, tech. rep., MIT, 1997.

[20] J. Yedidia, W. Freeman, and Y. Weiss, *Constructing Free Energy Approximations and Generalized Belief Propagation Algorithms*, IEEE Transactions on Information Theory, 51 (2005), pp. 2282–2312.