# Boundary conditions in generalized belief propagation, singularities of regionalizations

Daniel Bennequin

IMJ, Paris 7

**TGSI 2017**

CIRM 29 08 2017

*Joint work with Grégoire Sergeant-Perthuys, Olivier Peltre and Juan Pablo Vigneaux*

Ref. Information, énergies, messages, future arxiv

# 1 Context : finite statistical structures and learning

A family of finite sets $E_i; i \in I$, and a subset $A$ of $\mathcal{P}(I)$, seen as a poset with inclusion $\beta \subseteq \alpha$ denoted by $\alpha \to \beta$. Our main interests are a simplicial complex, specially a $PL$ manifold with boundary, or a network of disjoint cells $\alpha \in A$, connected by synapses $i \in I$. We denote by $E_\alpha$ the product of the $E_i$ where $i$ belong to $\alpha$, and $x_\alpha$ is the generic point of $E_\alpha$.
Data are *energy functions* $H_\alpha : E_\alpha \to \mathbb{R}; \alpha \in A$, giving weights $f_\alpha = e^{-\beta H_\alpha}$.
For this talk, we assume that $\forall \alpha, \forall x_\alpha, f_\alpha(x_\alpha) > 0$. We denote $f$ the product of all the $f_\alpha; \alpha \in A$.
Statistical Mechanics is interested with probability laws $p$ on the whole product $E = \prod_i E_i$.
The statistical equilibrium is the unique minimum of the *standard free energy* (Helmholtz, Boltzmann, Gibbs)

$$F_G(p) = E_p(-\beta^{-1} \ln f) - k_B^{-1} S(p). \tag{1}$$

Where $S(p) = -k_B \sum_x p(x) \ln p(x)$ is the entropy of $p$.
let us take $\beta = k_B = 1$ for simplicity. For this minimum $p^*$, $F_G(p^*) = -\ln Z$, the Helmholtz energy.

For a Bayesian learning models, we consider probabilities on a product $E \times \Theta$, where $\Theta$ is a finite set parameterizing probability laws $p_\theta; \theta \in \Theta$ on $E$, such that (tautologically) the conditional probability $p((x, \theta')|\theta' = \theta)$ is $p_\theta(x)$, and (by definition) the marginal over $x$ is the *a priori* law $p_a(\theta)$. Suppose that a subset $A'$ of $A$ is given, with a probability $p'$ on the product of the sets $E_{\alpha'}; \alpha' \in A'$ ; the aim of learning is to identify the probability $p_b$ on $\Theta$ which gives the better explanation of $p'$. It is given by the marginal $p_b$ of the law $P'$ on $E' \times \Theta$ which minimizes $F_V(P') = D_{KL}(P'; p' \otimes p_a)$, where $D_{KL}(P; Q) = \sum_y P(y) \ln \frac{P(y)}{Q(y)}$ is the Kullback-Leibler distance. A simple calculation shows that

$$F_V(P') = \mathbb{E}_{p_b}(-\ln p_a(\theta) + D_{KL}(X'_*(p_\theta); p') - S(p_b); \tag{2}$$

where $X'_*$ denotes the marginal on $E'$. Then we find the stat structure, with the log of *a priori* as an energy, and as we will see, $p'$ as a flux of information. However the new problem is how to compute $S(p_b)$ in a finite time ?

## 2 Generalized Bethe free energies, Kikuchi et al.

A *regionalization* $\mathcal{R}$ of the system is a subset of $\mathcal{P}(A)$, also seen as a poset, with $r \rightarrow s$ meaning $s \subseteq r$. We write $E_r = \prod_{i\in\alpha;\alpha\in r} E_i$, $f_r(x_r) = \prod_{\alpha\in r} f_\alpha(\pi_\alpha(x_r))$ and more generally, if $r \rightarrow s$, $f_{r,s}$,denotes the product of the $f_\alpha(\pi_\alpha(x_r))$, for $\alpha \subseteq r$, but $\alpha \nsubseteq s$.
$\mathcal{C}(\mathcal{R})$ denotes the family of probabilities on the sets $E_r$ for $r \in \mathcal{R}$, and $\mathcal{P}(\mathcal{R})$ the subset satisfying the gluing rules (of marginals) on all possible intersections of sets $r, s$ in $\mathcal{R}$ on the elements of $A$, a presheaf.
There exists a unique family of relative integers $(c_r), r \in \mathcal{R}$, such that

$$\forall r \in \mathcal{R}, \quad \sum_{t\in\mathcal{R}|t\rightarrow r} c_t = 1. \tag{3}$$

In fact,

$$c_r = \sum_{t|t\rightarrow r} \mu_{t,r}; \tag{4}$$

wher the Moebius function $\mu : \mathcal{R}\times\mathcal{R} \rightarrow \mathbb{Z}$ of the poset $\mathcal{R}$ satisfies

$$\forall r, s, \quad \delta_{r=s} = \sum_{t|r\rightarrow t\rightarrow s} \mu_{r,t} = \sum_{r\rightarrow t\rightarrow s} \mu_{t,s}. \tag{5}$$

**Definitions :** $F_{B,\mathcal{R}}(q) = \sum_{r\in\mathcal{R}} c_r F_r(q_r)$.
$\mathcal{R}^*$ is the subset of $\mathcal{R}$ defined by $c_r \neq 0$.
**Oral remarks and pictures :** boundaries and butterflies
The regions $r$ such that $c_r = 0$ appear as *singularities* in $\mathcal{R}$.
Definitions of $\mathcal{C}^*(\mathcal{R})$ and $\mathcal{P}^*(\mathcal{R})$ follow.
If we add the following constraints on $\mathcal{R}$ : $\sum_{r\in\mathcal{R}|i\in r} c_r = 1, \quad \sum_{r\in\mathcal{R}|\alpha\subset r} c_r = 1$;
we recover the usual internal energy, but only local entropies :

$$F_{B,\mathcal{R}}(q) = \mathbb{E}_q(-\ln f(x))) - \sum_{r\in\mathcal{R}} c_r S(q_r). \tag{6}$$

In the case of a $d$-dimensional $PL$ manifold $K$ with smooth boundary $\partial K$, and the simplicial subdivision, this gives

$$F_B(q) = \mathbb{E}_q(-\ln f) - \sum_{\alpha\in K^*} (-1)^{d-|\alpha|} S(q_\alpha), \tag{7}$$

$$F_B(q) = \mathbb{E}_q(-\ln f) - \sum_{\alpha\in K} (-1)^{|\alpha|+1} I_{|\alpha|}(X_\alpha; q_\alpha), \tag{8}$$

where $I_{|\alpha|}(X_\alpha; q_\alpha)$ denotes the mutual information of order $|\alpha|$ of all variables associated withh the vertices of $\alpha$.
With these signs the combination is "almost convex", cf. Baudot and Bennequin (announcement).
**Questions** : existence of a minimum ? unicity ? gluing in a total probability ?

## 3 Generalized belief propagation

A family of messages is a set of strictly positive functions $m_{r\to s}(x_s)$. Two families are considered to be equivalent if they differ by strictly positive multiplicative constants $\lambda_{r\to s}$. The propagation algorithm is

$$m'_{r\to s}(x_s) \approx \sum_{x_r \setminus x_s} f_{r,s}(x_r) \prod_{r'\to s' | r\to s', r\not\to r', s\not\to s'} m_{r'\to s'}(x_{s'}) \prod_{(r'\to s' | s\not\to r', r\to r', s\to s'} m^{-1}_{r'\to s'}(x_{s'}). \tag{9}$$

History: Gallager 63, Pearl 88, Yedidia, Freeman, Weiss 2001

Explanation : evolution of beliefs $b_r(x_r)$, that are probability laws :

$$n_{s\hookrightarrow r}(x_s) = \prod_{t\to s | r\not\to t} m_{t\to s}(x_s); \quad b_r(x_r) \approx \prod_{r\to s} n_{s\hookrightarrow r}(x_s); \tag{10}$$

$$\frac{m'_{r\to s}(x_s)}{m_{r\to s}(x_s)} \approx \frac{\sum_{x_r\setminus x_s} b_r(x_r)}{b_s(x_s)}. \tag{11}$$

We will see later a co-homological interpretation.

The Brouwer's fixed point theorem (plus a positivity argument) implies that there always exists fixed points of the algorithm.

Divide $\mathcal{R}\backslash\mathcal{R}^*$ in two parts $\mathcal{R}'$ and $\mathcal{R}$", such that there never exists $s' \to s$". And choose compatible probabilities $q^0_{R'_r}(x_{R'_r})$ over the subsets $\mathcal{R}'$ covered by the regions $r \in \mathcal{R}$ ; they are the *Dirichlet data*. And for each pair $r \to s$", with $s" \in \mathcal{R}"$ and $r \in \mathcal{R}^*$, a return message $n"_{s" \hookrightarrow r}(x_{s"})$, they are the *Neumann data*.

The modified algorithm is given by the preceding rules, except

$$\forall r \in \mathcal{R}, \quad b_r(x_r) = \frac{f_r(x_r)\prod_{s|r\to s} n_{s\mapsto r}(x_s) q^0_{\mathcal{R}'_r}(x_{\mathcal{R}'_r})}{\sum_{x_r\backslash x_{\mathcal{R}'_r}} f_r(x_r)\prod_{s|r\to s} n_{s\mapsto r}(x_s)}. \tag{12}$$

In this case also, there always exists fixed points of the algorithm.

$\mathcal{Q} = P^*(\mathcal{R};\mathcal{R}"|\mathcal{R}')$ is the set of probabilities on the $E_r; r \in \mathcal{R}$, that have coherent marginals at every $s$ in $\mathcal{R}\backslash\mathcal{R}"$, and induce the Dirichlet data in $\mathcal{R}"$.

Then consider the modified generalized Bethe (Kikuchi) function on $\mathcal{Q}$

$$F_{B,\mathcal{R},M}(q) = \sum_{r\in\mathcal{R}} c_r(\mathbb{E}_{q_r}(-\ln f_r(x_r) - \sum_{s"\hookrightarrow r} \ln n"_{s"\hookrightarrow r}(x_{s"})) - S(q_r)). \tag{13}$$

**Theorem** : Every interior critical point of $F_{B,\mathcal{R},M}$ is the belief of a fixed point of the modified generalized belief propagation algorithm $BPNGM$. Conversely, every (non-degenerate) fixed point of $BPNGM$ gives a critical belief.

This is a corrected version of the main result of Yedidia, Freeman and Weiss. See the key lemma below.
The usual implicit assumption is uniform homogeneous Neumann condition, i.e. all $n$" are equal to 1, then the fixed point satisfies $m_r = 1$ for each $r$ singular, in $\mathcal{R} \backslash \mathcal{R}^*$. However, in this case, $\mathcal{Q}$ is a set of probabilities in $R^*$ only, without gluing assumptions in $\mathcal{R} \backslash \mathcal{R}^*$.

# 4 Analogy and homology

Consider a domain $\Omega$ in $\mathbb{R}^N$, with a smooth boundary $\partial\Omega$, and the Poisson equation $\Delta u = f$. Two kinds of boundary conditions are usually considered : the Dirichlet condition, where the values of the solution are prescribed on the boundary $u(y) = q(y)$, and the Neumann condition, where the normal derivative along the boundary is given $\partial_\nu u(y) = n(y)$. They can be combined on different pieces $\partial'\Omega$ and $\partial"\Omega$ respectively of $\partial\Omega$.

The corresponding variational problem is

$$Min_u \int_\Omega (\frac{1}{2}|\nabla u(x)|^2 - f(x)u(x))dx - \int_{\partial"\Omega} u(y)n(y)d\sigma(y); \qquad (14)$$

where $u$ is constrained to be equal to given $q$ on $\partial'\Omega$.

It is tempting to consider $f$ as the analog of $H$, the kinetic energy as the analog of minus the entropy, the Dirichlet conditions as fixed beliefs, and the Neumann conditions as fixed incoming messages $n$".
Here too, homogeneous Neumann data correspond to unchanged functional and unchanged functional space.
Here too, singular points in the domain or on the boundary introduce difficulties.

In the case of a tree $\Gamma$, with roots and ending branches (picture), Pearl had shown in what sense fixed probability laws on ending branches are dual of incoming message in roots, giving a firm understanding of the belief propagation algorithm, from the principle of bayesian analysis.

With loops and higher dimensions, we enter in the field of Information Topology.

## A (small) part of the thesis of Olivier Peltre.

Here we suppose that $A$ contains the empty set $\emptyset$, and that for every $\alpha, \beta$ in $A$ the inresection $\alpha \cap \beta$ belongs to $A$. The height (resp. depth) of $\alpha$ is the maximal length of a non-degenerate chain starting (resp. ending) in $\alpha$. The notation $\mathfrak{g}^\alpha$ designate the vector space of the applications from $E_\alpha$ to $\mathbb{R}$ ; the space $\mathfrak{g}^\emptyset$ is reduced to 0. The canonical projections $\pi^{\alpha\beta} : E_\alpha \to E_\beta$ (where $\alpha \to \beta$) give injections $j_{\alpha\beta\beta} : \mathfrak{g}^\beta \to \mathfrak{g}^\alpha$, that make from $\mathfrak{g}$ a contravariant functor on $A$.

$A^{(n)}$ denote the set of non-degenerate chains of length $n$, and $\mathfrak{g}^{(n)}$ the direct sum of the spaces $g^{\underline{a}} \times 0_a$, where $\underline{a} = \partial_n a$ is the end of the chain $a$, when $a$ describe $A^{(n)}$. The *boundary operator* $\partial$ from $\mathfrak{g}^{(n)}$ to $\mathfrak{g}^{(n-1)}$ is defined by

$$\forall b \in A^{(n-1)}, \quad \partial_b(\varphi_a) = \sum_{k=0}^{n} (-1)^{n-k} \sum_{a' | \partial_k a' = b} \varphi_{a'}; \tag{15}$$

in such a way that $\partial \circ \partial = 0$.
In particular, from $\mathfrak{g}^{(2)}$ to $\mathfrak{g}^{(1)}$,

$$\forall \gamma \in A, \quad \partial_b(\varphi_{\alpha \to \beta}) = \sum_{\beta | \gamma \to \beta} \varphi_{\gamma \to \beta} - \sum_{\alpha | \alpha \to \gamma} \varphi_{\alpha \to \gamma}. \tag{16}$$

Olivier Peltre has computed the homology of $\partial$ and found that only $H_0$ is non-trivial and isomorphic with the convex-hull of the space $\mathcal{P}(X)$. The dimension is given explicitely by the Moebius function :

$$h_0(\partial) = \sum_{\alpha} \sum_{\beta} \mu_{\alpha\beta} |\beta|. \tag{17}$$

By using the marginal $j^*_{\alpha\beta}$, the exponential and logarithm, he defined a non-linear complex of dual operators.

**Definition** : the *non-linear co-boundary operator* from $g^{(1)}$ to $g^{(2)}$ is defined by

$$(D^*H)_{\alpha\to\beta}(x_\beta) = H_\beta(x_\beta) - \ln \sum_{x_\alpha \backslash x_\beta} e^{-H_\alpha(x_\alpha)}. \tag{18}$$

**Proposition** : the generalized belief propagation algorithm, at the level of $\ln b_\alpha$ consists to transform the collection $H_\alpha; \alpha \in A$ into the collection $H_\alpha + (\Delta H)_\alpha$, where

$$(\Delta H)_\alpha = \sum_{\beta|\alpha\to\beta} \partial_\beta (D^*H). \tag{19}$$

Then, the belief propagation algorithm can be written as a discrete non-linear heat equation :

$$\partial_\tau H = \Delta H. \tag{20}$$

**The key lemma** for establishing the theorem, by identification of Lagrange multipliers of the generalized and modified Bethe free energy with the log of the messages.

**Lemma** (adapted from Yedidia, Freeman, Weiss) : (i) the projection from $\mathcal{C}(\mathcal{R})$ to $\mathcal{C}^*(\mathcal{R})$ induces a bijection from $\mathcal{P}(\mathcal{R};\mathcal{R}_1)$ to $\mathcal{P}^*(\mathcal{R};\mathcal{R}_1)$.
(ii) The image of this projection is defined by the following "dual" set of equations :

$$\forall r \in \mathcal{R}, \forall s \in \mathcal{R}\backslash\mathcal{R}_1, r \to s, \forall x_s \in E_s : \sum_{t\in\mathcal{R}|t\to s, t\not\to r} c_t \sum_{x_t\backslash x_s} q_t(x_t) = 0. \quad (21)$$

This is true for any set of singular points $\mathcal{R}_1$, the case which is needed is $\mathcal{R}_1 = \mathcal{R}$".