2011 International Symposium on Nonlinear Theory and its Applications
NOLTA2011, Kobe, Japan, September 4-7, 2011

Graph zeta function and loopy belief propagation

Yusuke Watanabe† and Kenji Fukumizu†

†The Institute of Statistical Mathematics
10-3 Midoricho, Tachikawa, Tokyo 190-8562 Japan
Email: {watay, fukumizu}@ism.ac.jp

Abstract—This paper discusses a link between the loopy belief propagation (LBP) algorithm and the Graph zeta function. The LBP algorithm is a nonlinear iteration to approximate the marginal or posterior probabilities required for various statistical inference, using the graph structure to define the joint probability. The theoretical properties of the LBP algorithm are not easy to analyze because of the complex nonlinearity and the graph structure. The derived connection with Graph zeta function involves the mathematical relation with the properties of the graph, leading various theoretical analysis of the LBP algorithm.

1. Introduction

Graphical modeling is useful for describing a probability distribution using the local interactions among variables. It has been used in many fields such as statistics, statistical physics, coding theory, and so on. For statistical inference with a graphical model, we often need to marginal out some or all of the variables. This computation, however, causes combinatorial explosion for large graphs and discrete variables.

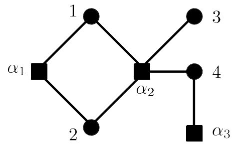
The belief propagation (BP, [8]) was first proposed as an efficient local propagation algorithm for computing all the marginals for trees. In later years, the iteration has been applied to graphs with cycles (Loopy BP (LBP) [6]), which has been found to give successful approximations in various applications, e.g. [4].

The theoretical properties such as convergence and correctness of the LBP algorithm, however, have not been clarified completely because of its nonlinear and combinatorial nature, though many works have been done ([14, 11, 3, 5] etc).

Following the previous work [13], this paper discusses a connection between the LBP algorithm and the graph zeta function. The main advances over [13] are (i) the connection is shown for a more general hypergraph (ii) the formulation by the Lagrange duality for the Bethe free energy is used.

2. Loopy BP and Bethe free energy

A *hypergraph* $H = (V, F)$ consists of a finite set of *vertices* V and a set of *hyperedges* F. A hyperedge is a non-empty subset of V, and often called a *factor*. For



$\alpha_1 = \{1,2\}$,
$\alpha_2 = \{1,2,3,4\}$,
$\alpha_3 = \{4\}$.

Figure 1: Bipartite graph representation.

$i \in V$, the *neighbors* of i is $N_i := \{\alpha \in F | i \in \alpha\}$, and for a hyperedge $\alpha \in F$ the neighbors of α is $N_\alpha := \{i \in V | i \in \alpha\} = \alpha$. $d_i := |N_i|$ and $d_\alpha := |N_\alpha| = |\alpha|$ are called degree. If the degree of every hyperedge is two, the hypergraph is identified with an ordinary graph. A hypergraph H can be represented as a bipartite graph with vertex nodes and factor nodes (Fig. 1)

A hypergraph H is *connected* (resp. *tree*) if the corresponding bipartite graph is connected (resp. tree). The nullity of H is denoted by $n(H)$; $n(H) = |V| + |F| - |\vec{E}|$. H is a tree iff it is connected and $n(H) = 0$.

Let $H = (V, F)$ be a hypergraph. For each $i \in V$, a variable x_i that takes values in $\mathcal{X}_i$ is associated. A probability density function p on $x = (x_i)_{i \in V}$ is said to be *factorized* with respect to H if it has the form

$$p(x) = \frac{1}{Z} \prod\nolimits_{\alpha \in F} \Psi_\alpha(x_\alpha), \tag{1}$$

where $x_\alpha = (x_i)_{i \in \alpha}$, Z is the normalization constant and Ψ_α are positive valued *compatibility functions.*

Many inference problems involve computation of marginal probaiblities. If the variables take a value in a finite set, a straightforward computational cost is exponential to the number of variables, thus infeasible for large graphs. In such cases, LBP approximates the marginals efficiently. It is known [15] that the LBP can be derived as a minimization procedure of Bethe free energy, which is used as an approximation of Gibbs free energy. We summarize their result below. Let $b_\alpha(x_\alpha)$ and $b_i(x_i)$ $(\alpha \in F, i \in V)$ be local probability density functions, which sum up to 1. The Bethe free energy is defined by

$$F(\{b_\alpha(x_\alpha), b_i(x_i)\}) = \sum_{\alpha \in F} \sum_{x_\alpha} b_\alpha(x_\alpha) \log \frac{b_\alpha(x_\alpha)}{\Psi_\alpha(x_\alpha)} + \sum_{i \in V} (1 - d_i) \sum_{x_i} b_i(x_i) \log b_i(x_i).$$

Consider the minimization problem on the restriction of "local consistency", i.e., $\sum_{x_{\alpha \setminus i}} b_\alpha(x_\alpha) = b_i(x_i)$ for each $i \in \alpha$. By introducing the Lagrangean multipliers, the condition of the stationary points can be expressed as [15]

$$m_{\alpha\to i}(x_i) = \omega \sum_{x_{\alpha \setminus i}} \Psi_\alpha(x_\alpha) \prod_{j\in\alpha, j\neq i} \prod_{\beta\ni j, \beta\neq\alpha} m_{\beta\to j}(x_j),$$

and

$$b_i(x_i) := \omega \prod_{\alpha\ni i} m_{\alpha\to i}(x_i) \quad (2)$$

$$b_\alpha(x_\alpha) := \omega \Psi_\alpha(x_\alpha) \prod_{j\in\alpha} \prod_{\beta\ni j, \beta\neq\alpha} m_{\beta\to j}(x_j), \quad (3)$$

where ω denotes (not necessarily the same) normalization constants. This gives a message passing method to obtain a stationary point of the minimization:

$$m^{t+1}_{\alpha\to i}(x_i) = \omega \sum_{x_{\alpha \setminus i}} \Psi_\alpha(x_\alpha) \prod_{j\in\alpha, j\neq i} \prod_{\beta\ni j, \beta\neq\alpha} m^t_{\beta\to j}(x_j), \quad (4)$$

We repeat Eq. (4) until the messages converge to a fixed point, though this is not guaranteed to converge. If it converges, we use Eqs.(2) and (3) for the approximations of the respective marginals of $p(x)$.

This paper discusses the case where the compatibility functions are expressed by exponential families. Let $\phi_i(x_i)$ and $\phi_{\langle\alpha\rangle}(x_\alpha)$ be statistics for a node variable x_i and a factor variable $x_\alpha = (x_j)_{j\in\alpha}$, respectively. We introduce a sufficient statistic ϕ_α by

$$\phi_\alpha(x_\alpha) = (\phi_{\langle\alpha\rangle}(x_\alpha), \phi_{i_1}(x_{i_1}), \ldots, \phi_{i_{d_\alpha}}(x_{i_{d_\alpha}})). \quad (5)$$

For describing the LBP algorithm, we prepare the exponential families for each factor α and each node i using the sufficient statistics ϕ_α and ϕ_i;

$$p_{\theta_\alpha}(x_\alpha) = \exp\left(\langle\theta_\alpha, \phi_\alpha(x_\alpha)\rangle - \psi_\alpha(\theta_\alpha)\right)$$

where $\theta_\alpha = (\theta_{\langle\alpha\rangle}; \theta_{\alpha:i_1}, \ldots, \theta_{\alpha:i_{d_\alpha}})$ is the natural parameter, and

$$p_{\theta_i}(x_i) \exp\left(\langle\theta_i, \phi_i(x_i)\rangle - \psi_i(\theta_i)\right)$$

with a natural parameter θ_i. Note that ϕ_α for a factor node includes the sufficient statistics for $i \in \alpha$ as its components in addition to $\phi_{\langle\alpha\rangle}$. The mean parameter $\eta_\alpha = (\eta_{\langle\alpha\rangle}, \eta_{\alpha:i_1}, \ldots, \eta_{\alpha:i_{d_\alpha}}) = \frac{\partial\psi_\alpha}{\partial\theta_\alpha}$ and $\eta_i = \frac{\partial\psi_i}{\partial\theta_i}$ serve also as parametrization for the exponential families.

Assume the compatibility functions have the form

$$\Psi_\alpha(x_\alpha) = \exp\left(\langle\bar{\theta}_\alpha, \phi_\alpha(x_\alpha)\rangle\right). \quad (6)$$

The following assumptions are indispensable to our analysis:

(A) For all $i \in V$ and $\alpha \in F$, the Hessian of ψ_i and ψ_α, (i.e. the covariance matrix) are invertible.

(B) the exponential families are "closed" under marginalization operation, i.e., for each pair of $i \in \alpha$, $\sum_{x_{\alpha \setminus i}} p_{\theta_\alpha}(x_\alpha)$ is included in p_{θ_i}.

It is not difficult to see that popular examples such as multinomial and Gaussian distributions satisfy these assumptions.

We rewrite the derivation of LBP from Bethe free energy in terms of exponential families. Suppose $b_\alpha(x_\alpha)$ and $b_i(x_i)$ have the form of exponential families. Then, the Bethe free energy is written by

$$F(\boldsymbol{\eta}) = -\sum_{\alpha\in F}\langle\bar{\theta}_\alpha, \eta_\alpha\rangle + \sum_{\alpha\in F}\varphi_\alpha(\eta_\alpha) + \sum_{i\in V}(1-d_i)\varphi_i(\eta_i),$$

where φ_α and φ_i are the negative entropy function for $p_\alpha(x_\alpha)$ and $p_i(x_i)$, respectively. The local consistency can be expressed by

$$\eta_{\alpha:i} = \eta_i \quad (i \in \alpha)$$

since this is equivalent to $\sum_{x_\alpha} \phi_i(x_i) b_\alpha(x_\alpha) = \phi_i(x_i) b_i(x_i)$.

This paper uses the Lagrange duality to derive the link to the graph zeta function. The Lagrange dual function G of F under the constraints of local consistency $\eta_{\alpha:i} = \eta_i$ ($\forall i \in \alpha$) is given by

$$G(\lambda) = \min_\eta F(\eta) + \sum_{i\in V}\sum_{\alpha\ni i}\lambda_{\alpha:i}(\eta_{\alpha:i} - \eta_i).$$

By taking the derivative, we see that the minimum is attained by $\theta_{\langle\alpha\rangle} = \bar{\theta}_{\langle\alpha\rangle}$, $\theta_{\alpha:i} = \bar{\theta}_{\alpha:i} - \lambda_{\alpha:i}$, and $\theta_i = \frac{1}{1-d_i}\sum_{\alpha\ni i}\lambda_{\alpha:i}$. Introduce $\mu_{\alpha:i}$ by

$$\mu_{\alpha\to i} := \lambda_{\alpha:i} - \frac{1}{d_i - 1}\sum_{\beta\ni i}\lambda_{\beta:i},$$

which has one-to-one correspondence with $(\lambda_{\alpha:i})$ by a linear map. In terms of $(\mu_{\alpha:i})$, the conditions of minimum are given by $\theta_{\langle\alpha\rangle} = \bar{\theta}_{\langle\alpha\rangle}$, $\theta_{\alpha:i} = \bar{\theta}_{\alpha:i} + \sum_{\beta\in N_i\setminus\alpha}\mu_{\beta\to i}$, and $\theta_i = \sum_{\beta\ni i}\mu_{\beta\to i}$. Let $h(\eta)$ denote the constraint function $h_{\alpha:i}(\eta) = \eta_{\alpha:i} - \eta_i$ ($i \in \alpha$), and $\eta_*(\lambda)$ be the point that attains the minimum. Then,

$$\frac{\partial F}{\partial\eta}(\eta_*(\lambda)) + \lambda^T\frac{\partial h(\eta_*(\lambda))}{\partial\eta} = 0$$

for all λ (or μ). Since the Lagrange dual function is given by $G(\lambda) = F(\eta_*(\lambda)) + \lambda^T h(\eta_*(\lambda))$, we have

$$\begin{aligned}\frac{\partial G}{\partial\lambda_{\alpha:i}} &= \Big(\frac{\partial F}{\partial\eta} + \lambda^T\frac{\partial h}{\partial\eta}\Big)\frac{\partial\eta}{\partial\lambda_{\alpha:i}} + h_{\alpha:i}(\eta_*(\lambda)) \\ &= \eta_{*\alpha:i}(\lambda) - \eta_{*i}(\lambda) \quad (7)\\ &= \Lambda_{\alpha:i}\Big(\bar{\theta}_{\langle\alpha\rangle}; \Big(\bar{\theta}_{\alpha:j} + \textstyle\sum_{\beta\in N_j\setminus\alpha}\mu_{\beta\to j}\Big)_{j\in N_\alpha}\Big) \\ &\quad - \Lambda_i\Big(\textstyle\sum_{\beta\in N_i}\mu_{\beta\to i}\Big), \quad (8)\end{aligned}$$

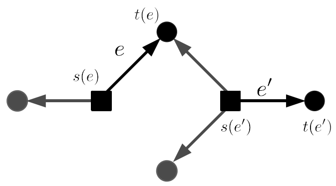


Figure 2: Example of the relation $e \rightharpoonup e'$.

where $\Lambda_{\alpha:i}(\theta_\alpha) = \eta_{\alpha:i}(\theta_\alpha)$ and $\Lambda_i(\theta_i) = \eta_i(\theta_i)$ are the mean maps.

It is not difficult to see the fixed point method for maximizing $G(\lambda)$,

$$\mu^{t+1}_{\alpha\to i} = \Lambda_i^{-1} \circ \Lambda_{\alpha:i}\Big(\overline{\theta}_{\langle\alpha\rangle}; \Big(\overline{\theta}_{\alpha:j} + \textstyle\sum_{\beta\in N_j\setminus\alpha}\mu^t_{\beta\to j}\Big)_{j\in N_\alpha}\Big) - \textstyle\sum_{\beta\in N_i\setminus\alpha}\mu^t_{\beta\to i}, \tag{9}$$

is equivalent to the BP update rule with $m^t_{\alpha\to i}(x_i) = \exp(\langle\mu^t_{\alpha\to i}, \phi_i(x_i)\rangle)$.

3. Graph zeta function

Ihara's graph zeta function [2] has been extended to arbitrary finite graphs ([10, 9]). This paper uses a further extension to hypergraphs with matrix weights, and connects it with the Hessian of the Bethe free energy, Let $H = (V, F)$ be a hypergraph. As noted before, it can be regarded as a bipartite graph. We use a directed graph notation by orienting each edge from a factor node α to a vertex node i. For each edge $e = (\alpha \to i)$, $s(e) = \alpha$ and $t(e) = i$ denote the starting factor and terminus vertex, respectively. If two edges $e, e' \in \vec{E}$ satisfy conditions $t(e) \in s(e')$ and $t(e) \neq t(e')$, this pair is denoted by $e \rightharpoonup e'$. (See Figure 2.) A sequence of directed edges $(e_1, \ldots, e_k)$ is said to be a *closed geodesic* if $e_l \rightharpoonup e_{l+1}$ for $l \in \mathbb{Z}/k\mathbb{Z}$. For a closed geodesic c, we may form the *m-multiple* c^m by repeating c m-times. A closed geodesic c is said to be *prime*, if it is not a multiple of strictly shorter closed geodesic. For example, a closed geodesic $c = (e_1, e_2, e_3, e_1, e_2, e_3)$ is not prime, while $c = (e_1, e_2, e_3, e_4, e_1, e_2, e_3)$ is prime. Two closed geodesics are *equivalent* if one is obtained by cyclic permutation of the other. An equivalence class of a prime closed geodesic is called a *prime cycle*. The set of prime cycles of H is denoted by $\mathfrak{P}_H$.

We associate each edge e with the size r_e of a matrix weight. The set of functions on $\vec{E}$ that take values on $\mathbb{C}^{r_e}$ for each $e \in \vec{E}$ is denoted by $\mathfrak{X}(\vec{E})$. The set of $n_1 \times n_2$ complex matrices is denoted by $\mathrm{M}(n_1, n_2)$.

Assume that for each $e' \rightharpoonup e$, a matrix weight $u_{e'\rightharpoonup e} \in \mathrm{M}(r_e, r_{e'})$ is associated. the (matrix weight) *graph zeta function* of H is defined by

$$\zeta_H(\boldsymbol{u}) := \textstyle\prod_{\mathfrak{p}\in\mathfrak{P}_H} \dfrac{1}{\det\big(I - \pi(\mathfrak{p})\big)},$$

where $\pi(\mathfrak{p}) := u_{e_k\rightharpoonup e_1} \ldots u_{e_2\rightharpoonup e_3} u_{e_1\rightharpoonup e_2}$ for $\mathfrak{p} = (e_1, \ldots, e_k)$.

The definition is an analogue to the Euler product formula of the Riemann zeta function which is represented by the product over all the prime numbers.

$\zeta_H(\boldsymbol{u}) = 1$ if H is a tree. For 1-cycle graph C_N of length N, the prime cycles are $(e_1, e_2, \ldots, e_N)$ and $(\bar{e}_N, \bar{e}_{N-1}, \ldots, \bar{e}_1)$. The zeta function is $\zeta_{C_N}(\boldsymbol{u}) = \det(I_{r_{e_1}} - u_{e_N\rightharpoonup e_1} \ldots u_{e_2\rightharpoonup e_3} u_{e_1\rightharpoonup e_2})^{-1} \det(I_{r_{\bar{e}_N}} - u_{\bar{e}_1\rightharpoonup \bar{e}_N} \ldots u_{\bar{e}_{N-1}\rightharpoonup \bar{e}_{N-2}} u_{\bar{e}_N\rightharpoonup \bar{e}_{N-1}})^{-1}$. Except for these two cases, the number of prime cycles is infinite.

Theorem 1 *Define a matrix $M(u)$ with index set $\mathfrak{X}(\vec{E})$ by*

$$\mathcal{M}(\boldsymbol{u})_{e,e'} = \begin{cases} u_{e'\rightharpoonup e} & \textit{if } e' \rightharpoonup e \\ 0 & \textit{otherwise.} \end{cases} \tag{10}$$

Then, the following formula holds

$$\zeta_G(\boldsymbol{u})^{-1} = \det(I - \mathcal{M}(\boldsymbol{u})). \tag{11}$$

For the proof, see [12]. This type of determinant formula is well known in the context of graph zeta functions (see e.g. [9], Theorem 3).

4. Bethe-Zeta formula

We have seen that the fixed point method for maximizing the dual function gives the LBP update. In this section, we will show that the Hessian of the Bethe free energy is related to the first derivative of the LBP update rule.

First, we consider the derivative of LBP update rule. From Eq.(9), we have

$$\frac{\partial \mu^{t+1}_{\alpha\to i}}{\partial \mu^t_{\beta\to j}} = \frac{\partial \theta_i}{\partial \eta_i}\frac{\partial \eta_{\alpha:i}}{\partial \theta_\alpha}\frac{\partial \theta_\alpha}{\partial \mu^t_{\beta\to j}} - (1 - \delta_{\alpha\beta})\delta_{ij}.$$

Note that only $\frac{\partial \theta_{\alpha:j}}{\partial \mu^t_{\beta\to j}}$-component for $\alpha \neq \beta$ in $\frac{\partial \theta_\alpha}{\partial \mu^t_{\beta\to j}}$ is nonzero. Thus the first term is $(1 - \delta_{\alpha\beta})\mathrm{Var}_{b_i}[\phi_i]^{-1}\mathrm{Cov}_{b_\alpha}[\phi_i, \phi_j]$. If the parameter satisfies local consistency, we see

$$\frac{\partial \mu^{t+1}_{\alpha\to i}}{\partial \mu^t_{\beta\to j}} = M_{\alpha\to i,\beta\to j}(u),$$

where $u = \mathrm{Var}_{b_i}[\phi_i]^{-1}\mathrm{Cov}_{b_\alpha}[\phi_j, \phi_i]$.

On the other hand, it follows from Eq.(7) that

$$\frac{\partial^2 G}{\partial \lambda_{\alpha:i}\partial \mu_{\beta\to j}} = \frac{\partial \eta_{\alpha:i}}{\partial \theta_\alpha}\frac{\partial \theta_\alpha}{\partial \mu_{\beta\to j}} - \frac{\partial \eta_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_{\beta\to j}}.$$

With the same u as above, it is also easy to see

$$(I - M(u))_{\alpha\to i,\beta\to j} = -\frac{\partial \theta_i}{\partial \eta_i}\frac{\partial^2 G}{\partial \lambda_{\alpha:i}\partial \mu_{\beta\to j}}.$$

Next, we derive a relation between the Hessian of F and G. Recall $\frac{\partial F}{\partial \eta} + \lambda^T \frac{\partial h}{\partial \eta} = 0$ for any $\eta = \eta_*(\lambda)$. By differentiating w.r.t. λ and noting h is a linear map, we have $\frac{\partial^2 F}{\partial \eta \partial \eta}\frac{\partial \eta}{\partial \lambda} + \frac{\partial \eta}{\partial \lambda} = 0$. Combining this with $\frac{\partial^2 G}{\partial \lambda \partial \lambda} = \frac{\partial h}{\partial \eta}\frac{\partial \eta}{\partial \lambda}$, we have

$$\frac{\partial^2 G}{\partial \lambda \partial \lambda} = -\frac{\partial \eta}{\partial \lambda}\frac{\partial^2 F}{\partial \eta \partial \eta}\frac{\partial \eta}{\partial \lambda}.$$

Based on the above facts, it is not difficult to derive the following theorem.

Theorem 2 *At any point $\boldsymbol{\eta} = \{\eta_{\langle\alpha\rangle}, \eta_i\}$ that satisfies local consistency, the following equality holds.*

$$\zeta_H(\boldsymbol{u})^{-1} = \det(I - \mathcal{M}(\boldsymbol{u}))$$
$$= \det(\nabla^2 F) \prod_{\alpha \in F} \det(\mathrm{Var}_{b_\alpha}[\phi_\alpha]) \prod_{i \in V} \det(\mathrm{Var}_{b_i}[\phi_i])^{1-d_i},$$

where

$$u^{\alpha}_{i \to j} := \mathrm{Var}_{b_j}[\phi_j]^{-1}\mathrm{Cov}_{b_\alpha}[\phi_j, \phi_i] \qquad (12)$$

is an $r_j \times r_i$ matrix, and $\nabla^2 F$ is the Hessian matrix with respect to the coordinate $\{\eta_{\langle\alpha\rangle}, \eta_i\}$.

A complete proof by a different approach is given in [12]. The above theorem tells that the determinant of Hessian of the Bethe free energy is given by the Graph zeta function at nonsymmetric correlation, which is related to the derivative of LBP update rule.

5. Applications to LBP

There are various results derived from the Bethe-Zeta formula. In this paper we focus only on convexity and positive-definiteness, and defer other results to [12] and a forthcoming paper.

The Bethe free energy function is not necessarily convex though it is an approximation of the Gibbs free energy function, which is convex. Non-convexity of the Bethe free energy can lead to multiple fixed points. [7] and [1] have derived sufficient conditions of the convexity and shown that the Bethe free energy is convex for trees and graphs with one cycle. In the following, L denotes the locally consistent beliefs $\{b_\alpha(x_\alpha), b_i(x_i)\}$, and $\|\cdot\|$ the maximum singular value.

Theorem 3 (Positive definite region)
Let κ be the Perron-Frobenius eigenvalue of $\mathcal{M}$, and define $L_{\kappa^{-1}} := \left\{\{b_\alpha(x_\alpha), b_i(x_i)\} \in L \mid \|\mathrm{Cor}_{b_\alpha}[\phi_i, \phi_j]\| < \kappa^{-1} \forall \alpha, i, j\right\}$. Then, the Hessian $\nabla^2 F$ is positive definite on $L_{\kappa^{-1}}(\mathcal{I})$.

Roughly speaking, as the degrees of factors and vertices increase, κ also increases and thus $L_{\kappa^{-1}}$ shrinks. The Perron-Frobenius eigenvalue is equal to 0 (resp. 1) if the hypergraph is a tree (resp. has a unique cycle). This result suggests that LBP works better for graphs of low degree.

The convexity of F depends solely on the given exponential family and the underlying hypergraph, because the Hessian $\nabla^2 F$ does not depend on the given compatibility functions, $\Psi = \{\Psi_\alpha\}$. For multinomial case, [7] have shown that the Bethe free energy function is convex if the hypergraph has at most one cycle. The following theorem extends the result. To show (ii), we need to capture the effect of cycles on arbitrary hypergraphs.

Theorem 4 *Let H be a connected hypergraph.*
(i) If $n(H) = 0$ or 1, then F is convex on L.
(ii) Assuming the exponential family is either a multinomial or Gaussian, then the converse of (i) holds.

References

[1] T. Heskes. On the uniqueness of loopy belief propagation fixed points. *Neural Computation*, 16(11):2379–2413, 2004.

[2] Y. Ihara. On discrete subgroups of the two by two projective linear group over p-adic fields. *J. Math. Soc. Japan*, 18(3):219–235, 1966.

[3] S. Ikeda, T. Tanaka, and S. Amari. Information geometry of turbo and low-density parity-check codes. *IEEE Trans. IT*, 50(6):1097–1114, 2004.

[4] R.J. McEliece and D.J.C.J.F. Cheng. Turbo decoding as an instance of Pearl's beliefpropagation algorithm. *IEEE J. Sel. Areas Commun.*, 16(2):140–52, 1998.

[5] J. M. Mooij and H. J. Kappen. Sufficient Conditions for Convergence of the Sum-Product Algorithm. *IEEE Trans. IT*, 53(12):4422–4437, 2007.

[6] K. Murphy, Y. Weiss, and M.I. Jordan. Loopy belief propagation for approximate inference: An empirical study. *Proc. UAI*, 15:467–475, 1999.

[7] P. Pakzad and V. Anantharam. Belief propagation and statistical physics. *Conf. Information Sciences and Systems*, 2002.

[8] J. Pearl. Reverend Bayes on inference engines: A distributed hierarchical approach. In *Proc. AAAI National Conf. AI*, pages 133–136, 1982.

[9] H.M. Stark and A.A. Terras. Zeta functions of finite graphs and coverings. *Adv. in Math.*, 121(1):124–165, 1996.

[10] T. Sunada. L-functions in geometry and some applications. *Lecture Notes in Math*, 1201:266–284, 1986.

[11] M.J. Wainwright, T.S. Jaakkola, and A.S. Willsky. Tree-based reparameterization framework for analysis of sum-product and related algorithms. *IEEE Trans. IT*, 49(5):1120–1146, 2003.

[12] W. Watanabe and K. Fukumizu. Loopy belief propagation, bethe free energy and graph zeta function. *Submitted.*, 2011.

[13] Y. Watanabe and K. Fukumizu. Graph zeta function in the bethe free energy and loopy belief propagation. In *Adv. in NIPS* 22, 2017–2025. MIT Press, 2010.

[14] Y. Weiss. Correctness of Local Probability Propagation in Graphical Models with Loops. *Neural Computation*, 12(1):1–41, 2000.

[15] J.S. Yedidia, W.T. Freeman, and Y. Weiss. Generalized belief propagation. *Adv. in NIPS*, 13:689–95, 2001.