J. Phys. A: Math. Gen. **LETTER TO THE EDITOR**

Information geometry and Plefka's mean-field theory

C. Bhattacharyya† § and S. Sathiya Keerthi‡

† Dept. of Computer Science and Automation. Indian Institute of Science, Bangalore, India

‡ Dept. of Mechanical and Production Engineering, National university of Singapore, Singapore

Abstract. An alternate derivation of Plefka's method is presented which is obtained from minimizing Gibbs energy. It is also shown that the method can be rederived using a perturbation expansion of Kullback Leibler divergence, the latter having a nice information geometric interpretation.

1. Introduction

Plefka[2] suggested an alternative derivation of Thouless, Anderson, and Palmer (TAP) equations for the Sherrington-Kirkpatrick (SK) model. He showed that a power expansion of the Gibbs energy up to the second order in exchange couplings for the SK model yields the TAP equations. This method is an interesting alternative to the diagram expansion method presented in [3]. In this paper we study Plefka's method and establish that this method can also be derived from a minimization framework, and also from an information geometric viewpoint.

2. An alternate derivation of Plefka's Mean-field theory

Plefka's method was initially presented for the SK model whose Hamiltonian, described by N Ising spins ($S_i = \pm 1$), is given by

$$\mathcal{H} = -\frac{1}{2}\sum_{i \neq j} J_{ij} S_i S_j \tag{1}$$

The interactions, J_{ij}, are identical and independently distributed random variables given by a gaussian distribution. Plefka's method actually holds for a general Hamiltonian, $\mathcal{H}$. We will thus try to study a system with N interacting spins, whose Hamiltonian, $\mathcal{H}$ is a general function of S.

§ To whom correspondence should be addressed.

Let γ be a real parameter taking values from 0 to 1. It is used to define the γ dependent Hamiltonian, $\gamma\mathcal{H}$. The probability distributions associated with the Hamiltonians H and $\gamma\mathcal{H}$ are given by:

$$p_\gamma = e^{-\beta\gamma\mathcal{H}-\phi_\gamma}\ , \quad p = e^{-\beta\mathcal{H}-\phi} \tag{2}$$

where $\phi_\gamma = \ln\sum_{\{S_i\}} e^{-\beta\gamma\mathcal{H}}$, $\phi = \ln\sum_{\{S_i\}} e^{-\beta\mathcal{H}}$ and $\beta = (KT)^{-1}$. At $\gamma = 1$ it is noted that $p = p_\gamma$. We introduce a set of external magnetic fields $\{h_i^{ex}\}$ to rewrite ϕ_γ as

$$\phi_\gamma = \ln\sum_{\{S_i\}} e^{-\beta\left(\gamma\mathcal{H}-\sum_i h_i^{ex}S_i+\sum_i h_i^{ex}S_i\right)+\tilde{\phi}_\gamma-\tilde{\phi}_\gamma} \tag{3}$$

where $\tilde{\phi}_\gamma = \ln\sum_{\{S_i\}} e^{-\beta(\gamma\mathcal{H}-\sum_i h_i^{ex}S_i)}$. We introduce one more Hamiltonian $\tilde{\mathcal{H}}$ and distribution $\tilde{p}_\gamma$,

$$\tilde{\mathcal{H}} = \gamma\mathcal{H} - \sum_i h_i^{ex}S_i\ ,\ \tilde{p}_\gamma = e^{-\beta\tilde{\mathcal{H}}-\tilde{\phi}_\gamma} \tag{4}$$

Applying the convex inequality $\langle e^x\rangle \geq e^{\langle x\rangle}$, to equation (3), and using equation (4) the following inequality is obtained,

$$-\phi_\gamma \leq -\tilde{\phi}_\gamma + \beta\sum_i h_i^{ex}u_i \tag{5}$$

where

$$u_i = \langle S_i\rangle_{\tilde{p}_\gamma} = \frac{1}{\beta}\frac{\partial\tilde{\phi}_\gamma}{\partial h_i^{ex}} \tag{6}$$

The inequality yields an upper bound G on $-\phi_\gamma$. The upper bound is tight if we set

$$h_i^{ex} = 0\ \forall\ i \tag{7}$$

This solution in no way helps us in estimating the means, u_i, of the system. It is thus appropriate to consider the right hand side of the inequality (5) as a function of $\{u_i\}$. This is achieved by treating h_i^{ex} as being dependent on $\{u_i\}$ via the following assumption,

Assumption. For a given $\{u_i\}$, γ, β one can solve for $\{h_i^{ex}\}$ in (6) uniquely.

We can thus define

$$\beta G(\gamma,\beta,\{u_i\}) = -\tilde{\phi}_\gamma + \beta\sum_i h_i^{ex}u_i \tag{8}$$

where G is the Gibbs free energy for the Hamiltonian $\tilde{\mathcal{H}}$. The relation (5) can be restated as

$$-\phi_\gamma \leq \beta G(\gamma,\beta,\{u_i\}) \tag{9}$$

At a given β and γ tightening the bound corresponds to minimization of the convex function G with respect to $\{u_i\}$. In fact at the minimum point the bound is tight and exact equality is attained. The above statements are established by examining the stationarity conditions

$$\frac{\partial G}{\partial u_i} = 0 \tag{10}$$

Since $\frac{\partial G}{\partial u_i} = h_i^{ex}$, this equation is a restatement of (7); and using (8), (9) we obtain:

$$-\phi_\gamma = \min_{\{u_i\}} \beta G(\gamma, \beta, \{u_i\}) \tag{11}$$

We establish the convexity of G with respect to $\{u_i\}$ by showing that the Hessian is positive definite. Let $\mathbf{H}$ denote the Hessian of G:

$$\mathbf{H_{ij}} = \frac{\partial^2 G}{\partial u_i \partial u_j} \tag{12}$$

Noting the fact that $\mathbf{H}_{ij} = \frac{\partial h_i^{ex}}{\partial u_j}$ and differentiating (6) with respect to $\{u_i\}$ we obtain

$$\mathbf{I} = \mathbf{BH} \tag{13}$$

where

$$\mathbf{B}_{ij} = \frac{1}{\beta^2} \frac{\partial^2 \tilde{\phi}_\gamma}{\partial h_j^{ex} \partial h_i^{ex}} = \frac{1}{\beta} \frac{\partial u_i}{\partial h_j^{ex}} = [\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle] \tag{14}$$

Being a covariance matrix $\mathbf{B}$ is a positive semidefinite matrix. Equation (13) ensures that both $\mathbf{B}$ and $\mathbf{H}$ are nonsingular. Hence $\mathbf{H}$ is a positive definite matrix; clearly then, G is convex.

The use of (10) as the underlying mean-field equations, so as to satisfy (7), and βG as an approximation of $-\phi_\gamma$ are the basic aspects of Plefka's theory. The Legendre transformation technique employed in [2] elegantly combines the invertibility assumption, with the fact that $\frac{\partial G}{\partial u_i} = h_i^{ex}$ and $\frac{\partial \tilde{\phi}_\gamma}{\partial h_i^{ex}} = \beta u_i$. But in the Legendre transformation approach, the need for defining G arises because of convenience in manipulation, whereas in our derivation G appears much more naturally. More importantly our analysis establishes the relation (9), which is not at all obvious from Plefka's derivation. Consequently, (10) comes from a variational argument (the convexity of G lends excellent support to this), in our derivation. Plefka's method can thus be seen as implementing minimization of Gibbs energy, over the set of all systems whose Hamiltonian are parametrized by (4).

Unfortunately the intractability in inverting (6) for $\gamma \neq 0$, makes it impossible to arrive at an algebraic expression for G at $\gamma = 1$. To circumvent this problem an approximate description of G is built, by using Taylor series expansion around $\gamma = 0$. Suppressing the dependence of β, we obtain

$$\tilde{G}_M(\gamma, \{u_i\}) = G(\{u_i\}, 0) + \sum_{k=1}^{M} \frac{\gamma^k}{k!} \frac{\partial^k G}{\partial \gamma^k}\bigg|_{\gamma=0} \tag{15}$$

The requirement mentioned in (7) is approximately enforced by

$$h_i^{ex} = \frac{\partial G}{\partial u_i} \approx \frac{\partial \tilde{G}_M}{\partial u_i} = 0 \tag{16}$$

This equation is used to set up the fixed point equations. For $M = 1$ we obtain the standard mean-field equations, while the TAP equations, for the SK model, are obtained by setting $M = 2$. Henceforth we will refer them as the mean-field equations.

Apart from establishing a minimization framework the relation (9) can also be interpreted using the Kullback-Leibler (KL) divergence between two distributions p and q, defined by

$$D(p, q) = \sum_{\{S_i\}} p \ln \frac{p}{q} \tag{17}$$

D is non-negative, and $D(p, q) = 0$ if and only if $p = q$. Non-negativity of $D(\tilde{p}_\gamma, p_\gamma)$ can be used to give an alternative route for arriving at the inequality (9). Minimizing this divergence with respect to the variational parameters, $\{u_i\}$, is equivalent to minimizing the convex function G. At the minimum point the divergence is zero. This observation leads to a information geometric interpretation presented in the next section.

3. A geometric argument

Mean-field theory can also be described from an information geometric viewpoint, developed by Amari et. al [1]. This geometric interpretation suggests that Plefka's method can also be derived from a Taylor series approximation of the divergence measure.

Let $\mathcal{M} = \{q(\{S_i\})\}$ be the set of all probability distributions, over the state space $\mathcal{S}$ consisting of 2^N states of $\vec{S}$. Since $\sum_{\{S_i\}} q(\{S_i\}) = 1$ and $0 < q < 1$, $\mathcal{M}$ can be thought of as a manifold of dimension $2^N - 1$ each of whose points represents a Boltzmann distribution $q(\{S_i\}) = e^{-\beta E - \phi}$, $\phi = \ln \sum_{\{S_i\}} e^{-\beta E}$. Consider the energy function in (4) and the corresponding probability distribution defined in (2). Let us define the following subsets of $\mathcal{M}$: $\mathcal{A}(\gamma) = \{\tilde{p}_\gamma | \gamma\}$, $\mathcal{B}(\{u_i\}) = \left\{p_\gamma | \langle S_i \rangle_{\tilde{p}_\gamma} = u_i\right\}$. $\mathcal{A}(\gamma)$ is a sub manifold with a fixed γ which is parametrized by $\{h_i^{ex}\}$. $\mathcal{A}(0)$ consists of points that correspond to factorial distributions; henceforth we will refer to $\mathcal{A}(0)$ as the factorized sub manifold. $\mathcal{B}(\{u_i\})$ is also a sub manifold where the mean vector, $\{u_i\}$, is fixed; $\mathcal{B}(\{u_i\})$ is a one dimensional sub manifold parametrized by γ. By the invertibilty assumption we can say that at each point of $\mathcal{A}(\gamma)$ there exists a unique $\{u_i\}$ such that $\mathcal{B}(\{u_i\})$ passes through that point. In fact they intersect one another orthogonally; see (13) and [1]. **B** plays the role of Riemannian metric on the manifold $\mathcal{M}$ (see [1] for details). Let us now consider the following problem: compute

$$\langle S_i \rangle_{p_\gamma} = \sum_{\vec{S}} S_i p_\gamma$$

for a fixed γ where p_γ is as in (2).

The divergence measure in the manifold $\mathcal{M}$ is given by the Kullback Leibler divergence defined in (17)[1]. Select any point $\tilde{p}_\gamma$ in $\mathcal{A}(\gamma)$. Then

$$D(\tilde{p}_\gamma, p_\gamma) = \sum_{\{S_i\}} \tilde{p}_\gamma \ln \frac{\tilde{p}_\gamma}{p_\gamma} = \phi_\gamma - \tilde{\phi}_\gamma + \beta \sum_i h_i^{ex} u_i \tag{18}$$

We immediately obtain the relation (5). Minimizing this divergence with respect to $\{h_i^{ex}\}$ trivially finds the solution $\{h_i^{ex} = 0\}$. Since the quantity of interest is $\{u_i\}$ and the above solution in no way helps in evaluating it we try to express D in (18) as a function of $\{u_i\}$. As discussed before, there exists a unique sub manifold $\mathcal{B}(\{u_i\})$ passing through the selected point $\tilde{p}_\gamma$ which helps in expressing D as a function of $\{u_i\}$:

$$D(\tilde{p}_\gamma, p_\gamma) = \beta G(\{u_i\}, \gamma) + \phi_\gamma \tag{19}$$

where G is as defined in (8). The mean-field criterion can now be stated as

$$\min_{\{u_i\}} D(\tilde{p}_\gamma, p_\gamma) \tag{20}$$

This minimization leads us to solving the equations (10). Since we cannot express $\frac{\partial G}{\partial u_i}$ as an explicit function of $\{u_i\}$, as discussed before we resort to an approximate description of G as a function of $\{u_i\}$ using Taylor series expansion. For an explicit description of G we turn to the factorized sub manifold, $\mathcal{A}(0)$, where we have an algebraic expression for $\{h_i^{ex}\}$ as a function of $\{u_i\}$. Let $\tilde{p}_0$ be the unique point, $\mathcal{A}(0) \bigcap \mathcal{B}(\{u_i\})$. The following expressions for distances can be easily verified:

$$D(\tilde{p}_0, \tilde{p}_\gamma) = \beta \tilde{G}_1 - \beta G \quad , \quad D(\tilde{p}_0, p_\gamma) = \beta \tilde{G}_1 + \phi_\gamma \tag{21}$$

where $\tilde{G}_1$ is obtained by setting $M = 1$ in (15). It directly follows from (19) and (21) that

$$D(\tilde{p}_0, p_\gamma) = D(\tilde{p}_0, \tilde{p}_\gamma) + D(\tilde{p}_\gamma, p_\gamma) \tag{22}$$

This pythagorean relationship is used to obtain an alternate definition of $D(\tilde{p}_\gamma, p_\gamma)$, and hence an alternate mean-field criterion,

$$\min_{\{u_i\}} D(\tilde{p}_\gamma, p_\gamma) = \min_{\{u_i\}} \{D(\tilde{p}_0, p_\gamma) - D(\tilde{p}_0, \tilde{p}_\gamma)\} \tag{23}$$

$D(\tilde{p}_0, \tilde{p}_\gamma)$ is still intractable, but we can build an approximate description in terms of $\{u_i\}$, and γ, by Taylor series expansion in γ around $\gamma = 0$:

$$D(\tilde{p}_0, \tilde{p}_\gamma) = D(\tilde{p}_0, \tilde{p}_0) + \gamma \left. \frac{\partial D}{\partial \gamma} \right|_{\gamma=0} + \sum_{k=2}^{M} \gamma^k \left. \frac{\partial^k D}{\partial \gamma^k} \right|_{\gamma=0} \tag{24}$$

The first two terms of right hand side are zero while the remaining terms are the negative of partial derivatives of G with respect to γ at $\gamma = 0$, which are all tractable. Note that since our operation is restricted to $\mathcal{B}(\{u_i\})$, $\{u_i\}$ is fixed.

Hence the mean-field criterion, (23), can be restated as

$$\min_{\{u_i\}} D(\tilde{p}_\gamma, p_\gamma) \approx \min_{\{u_i\}} \beta\tilde{G}_M(\{u_i\},\gamma) + \phi_\gamma \tag{25}$$

where $\tilde{G}_M$ is defined in (15). We obtain our mean-field equations, (16), by solving for stationarity conditions.

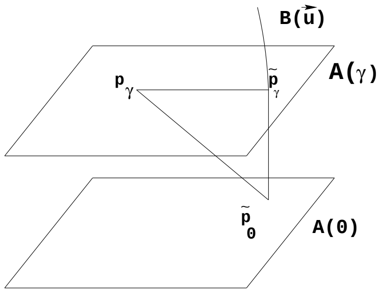

Figure 1. A pictorial illustration of $\mathrm{D}(\tilde{p}_\gamma, p_\gamma) = D(\tilde{p}_0, p_\gamma) - D(\tilde{p}_0, \tilde{p}_\gamma)$. The first order approach corresponds to approximating the divergence between p_γ and $\tilde{p_\gamma}$ by that between p_γ and $\tilde{p_0}$. The divergence between $\tilde{p_\gamma}$ and $\tilde{p_0}$ consists of second and higher order terms in γ. Subtracting these terms from the divergence between p_γ and $\tilde{p_0}$ yields a better approximation of the divergence between p_γ and $\tilde{p_\gamma}$. This corresponds to Plefka's approach.

It is easy to see from (21) - (22) that when p_γ is not factorial, $\min_{\{u_i\}} D(\tilde{p_0}, p_\gamma)$ is positive, and hence cannot yield an arbitrary close approximation to ϕ. It is also straightforward to establish that the objective function G is overestimated by $\tilde{G}_1$. Note using (19), (21) and the non-negativity of D that

$$-\phi_\gamma \leq \beta G(\{u_i\},\gamma) \leq \beta\tilde{G}_1(\{u_i\},\gamma).$$

These interesting inequalities are, again, not at all obvious in Plefka's derivation.

Thus Plefka's method can also be derived from an perturbation expansion of the KL divergence. The derivation of basic ideas of mean-field theory done in this section is an elegant geometric alternative to the algebraic derivation presented in the previous section.

4. Discussion

In this paper we have studied Plefka's method and have established that it can also be derived from minimization of Gibbs energy framework. It is also estabished that this method can be obtained from a perturbation expansion of KL divergence, which yields an information geometric interpretation.

References

[1] Amari, S., Kurata, K., Nagaoka, H.(1992) Information geometry of Boltzmann Machines. *IEEE Transactions on Neural Networks, Vol 3, No 2*,260-271.

[2] Plefka, T. Convergence condition of the TAP equation for the Infinite-ranged Ising glass model. *J. Phys. A: Math. Gen.*15(6) 1971-1978.

[3] Thouless, D. J., Anderson, P. W., Palmer, R. G. (1977) Solution of 'Solvable model of a spin glass'. *Phil. Mag. 35*(3) 593-601.