0.1. Graphical models as Markov Random Fields

One of the basic concepts in statistical modelling are graphical models which greatly help in analysing multivariate phenomenon. Visualization by graphs help with efficient development and understanding analysed models while complex computations can be performed Consider a graph $G = (V, E)$ which consists of a finite set of vertices V and a collection of edges $E \subset V \times V$. Each edge $e_i \in E$ joins two vertices and in general may have a direction. The vertex $v \in V$ may be seen as a random variable X_v defined on some space $\mathcal{X}_v$ that may be either continuous or discrete. We will use the notion of *clique* which is a subset of V in which all nodes are pairwise connected. One of the most useful family of models are Markov Random Fields which are type undirected random fields which satisfies global Markov property, specifically:

Definition 1 *[koller2009probabilistic] An undirected graphical model G is a Markov Random Field (MRF) if for any node X_v in the graph the following conditional property holds:*

$$P(X_i|X_{G\setminus i}) = P(X_i|X_{N(i)})$$

where $X_{G\setminus i}$ denotes all the nodes except X_i, and $X_{N(i)}$ denotes the set of all vertices connected to X_i.

Thus, the MRF has a desired property that any two nodes are conditionally independent given some evidence nodes that separates them. This property is closely related with the notion of factorization of the joint probability distribution:

Definition 2 *A probability distribution $P(athbf{X})$ on an undirected graphical model G factorizes over G if there exists a set of non-negative functions (potentials) $\{\psi_C\}_{C \in \mathcal{C}}$ functions on cliques that cover all the nodes and edges of G and we can write:*

$$P(X_1, X_2, ..., X_n) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C)$$

where $\mathcal{C}$ is a set of all cliques in G and Z is a normalization constant $Z = \sum_{\mathbf{x}} \prod_{C \in \mathcal{C}} \psi_C(X_C)$ which is often called a partition function.

Theorem 1 (Hammersley-Clifford, 1972?) *Strictly positive distribution $P(\mathbf{X})$ is MRF w.r.t an undirected graph G if and only if it factorizes over G.*

This theorem ensures us that there exits a general factorization form of the distribution of MRFs. It follows from the strict positivity of P that we can write:

$$p(x_1, x_2, ..., x_n) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C) = \frac{1}{Z} e^{\sum_{C \in \mathcal{C}} \ln \psi_C(x_C)} = \frac{1}{Z} e^{-E(x)}$$

where $E(x)$ is called an energy function. This general form of distribution is usually defined as *Gibbs distribution*. Hence, the probability distribution of every MRF can be expressed as in **??**. This relationship allow us to take advantage of Moreover, this form of distribution is a natural candidate to approximate and model phenomenon which can be seen as graphical models. In next sections we will analyse one of examples of Gibbs which will is powerful enough to approximate any probability distribution (TODO - find concept of apprximation by BM)

0.2. Boltzmann distribution

In this thesis, a primary undirected graphical model (or MRF) which will be analysed in the general form has the joint distribution:

$$p(x_1, x_2, ..., x_n) = \frac{1}{Z} \exp(-E(x_1, x_2, ..., x_n))$$

where E is the *energy* of the system of the form:

$$E(\mathbf{X}) = -\sum_{(ij)} w_{ij} x_i x_j - \sum_i \theta_i x_i$$

The pair-wise potential function have here the form:

$$\psi_{i,j} = \exp(x_i w_{ij} x_j)$$

while the magnetic fields defined as:

$$\psi_i = \exp(a_i x_i)$$

The name comes from the Boltzmann distribution which extensively used in physics to compute the energy of the system of particles. This model proves to be very useful for many applications such as the error-correcting code, computer vision, medical diagnosis or statistical mechanics [**yedidia2001idiosyncratic**]. This model may represents the statistical dependencies between different variables through the weight link w_{ij} as well as the evidence for the specific variable. However, computing the partition function requires to sum over a number of states that grows exponentially with the number of variables and is untractable even for a small number of variables. That is why, we have to resort to some tractable approximation which one of them will be analysed in next section.

0.3. Statistical Perspective

Consider energy based model (initially without latent variables) of the form:

$$P(\mathbf{s}) = \frac{e^{-E(\mathbf{s})}}{\sum_{\mathbf{s}} e^{-E(\mathbf{s})}} = \frac{1}{Z} e^{-E(\mathbf{s})} \tag{1}$$

where energy is defined as:

$$E \equiv E(\mathbf{s}) = -\sum_{(ij)} s_i w_{ij} s_j - \sum_i a_i s_i$$

If we restrict each node to have two states, $s_i \in \{0, 1\}$, we obtain well-known the Ising model [**yedidia2001idiosyncratic**]. Restricting the w_{ij} to be positive we obtain the ferromagnetic Ising model. Finally, assuming that the w_{ij} are chosen from a random distribution, we obtain the Ising spin glass model. Instead of imposing some restrictions on model structure we will try to find a approximate distribution Q (which poses useful characteristics) that minimizes relative entropy often called Kullback-Leibler divergence:

$$KL(Q||P) = \mathbb{E}_Q \left(\ln \frac{Q}{P} \right) = \sum_{\mathbf{s}} Q(\mathbf{s}) \ln \frac{Q(\mathbf{s})}{P(\mathbf{s})} \tag{2}$$

The KL-divergence is non-symmetric measure of the difference between two distributions which is always non-negative. Substituting P from 1 into previous equation gives us:

$$KL(Q||P) = \ln Z + E[Q] - H[Q]$$

where H stands for entropy of the distribution Q, $\ln Z$ is the *free energy* and $E[Q] = \sum_{\mathbf{s}} Q(\mathbf{s}) E(\mathbf{s})$ is called the *variational energy* [**opper2001advanced**]. The partition function Z doesn't depend on Q and we need to only focus on minimizing the variational free energy:

$$F[Q] = E[Q] - H[Q]. \tag{3}$$

0.4. Mean Field Approximation

0.5. Extended Mean Field Approximation

Mean field approach (TODO: murphy here why MF):

$$Q(\mathbf{s}) = \prod_i q_i(x_i) \tag{4}$$

MF theory is exact for infinite-ranged Ising model. In the case of spin glass model Thouless, Anderson and Palmer (TAP) showed the corresponding Gibbs free energy with additional "Onsager term".

$$\ln Z = \ln \sum_{\mathbf{s}} \exp(-E(\mathbf{s})) = \ln \sum_{\mathbf{s}} q(\mathbf{s}) \frac{\exp(-E(\mathbf{s}))}{q(\mathbf{s})} \geqslant \sum_{\mathbf{s}} q(\mathbf{s}) \ln \frac{\exp(-E(\mathbf{s}))}{q(\mathbf{s})} = -\mathbb{E}(E(\mathbf{s})) + H(Q) = -F \tag{5}$$

The bound F on log Z is called *the mean field free energy.*

We will minimize 3 where we instead of assuming Q to be a product distribution we require that our distribution has to satisfy:

$$\mathbb{E}_Q(\mathbf{S}) = \mathbf{m} \tag{6}$$

where $\mathbb{E}$ refers to the average configuration under the Boltzmann measure). Thus, the Gibbs' free energy is defined as

$$G(\mathbf{m}) = \min_Q \{E(Q) - H(Q) \mid \mathbb{E}(\mathbf{S}) = \mathbf{m}\} \tag{7}$$

The constrained optimization problem can be transformed into unconstrained using Lagrange multipliers, i.e.:

$$E(Q) - H(Q) - \sum_i \lambda_i (S_i - m_i) \tag{8}$$

which leads to the new form of the free energy with auxiliary fields:

$$F(\boldsymbol{\lambda}) = \sum_{\mathbf{s}} \exp(-E(\mathbf{s}) - \sum_i \lambda_i s_i) \tag{9}$$

Using this new free energy in the equation 7 yields:

$$G(\mathbf{m}, \boldsymbol{\lambda}) = \sum_i \lambda_i m_i - \ln \sum_{\mathbf{s}} \exp(-E(\mathbf{s}) + \sum_i \lambda_i s_i) \tag{10}$$

Through the Legendre transform we can conjugate both unknown parameters and the condition 6 on the λ_i can be introduced.

$$G(\mathbf{m}) = \max_{\boldsymbol{\lambda}} \{\sum_i \lambda_i m_i - \ln \sum_{\mathbf{s}} \exp(-E(\mathbf{s}) + \sum_i \lambda_i s_i)\} \tag{11}$$

where now the the maximizing auxiliary field $\boldsymbol{\lambda}^*(\mathbf{m})$ is the inverse function of $\mathbf{m}(\boldsymbol{\lambda}) = \frac{\mathrm{d}F}{\mathrm{d}\boldsymbol{\lambda}}$. As we operate on the concave function, the Legendre transform is its own inverse and and we can restore the free energy by setting the auxiliary field to zero which yields:

$$F = \min_{\mathbf{m}} G(\mathbf{m}). \tag{12}$$

Now lets assume inifinite temperature. - expand there Here Georges: Variational free energy with a set of external auxiliary fields (Lagrange multipliers):

$$-\beta G = \ln \sum_{\mathbf{s}} \exp\left(\beta \sum_{(ij)} w_{ij} s_i s_j + \beta \sum_i a_i s_i + \sum_i \lambda_i(\beta)(s_i - m_i)\right)$$

We are interested in expanding $-\beta G(\beta, \mathbf{m})$ around $\beta = 0$ where the spins are entirely controlled by their auxiliary fields:

$$-\beta G = -(\beta G)_{\beta=0} - \left(\frac{\partial(\beta G)}{\partial \beta}\right)_{\beta=0} \beta - \left(\frac{\partial^2(\beta G)}{\partial \beta^2}\right)_{\beta=0} \frac{\beta^2}{2} - \dots$$

At $\beta = 0$ the spins are uncorrelated:

$$m_i = \mathbb{E}_{\beta=0}(s_i) = \frac{\exp(\lambda_i(0))}{\exp(\lambda_i(0)) + 1} = \text{sigmoid}(\lambda_i(0)) \tag{13}$$

and

$$\begin{aligned}
-(\beta G)_{\beta=0} &= \ln \sum_{\mathbf{s}} \exp\left(\sum_i \lambda_i(0)(s_i - m_i)\right) \\
&= \ln \left\{\sum_{s_1} \exp\left(\lambda_1(0)(s_1 - m_1)\right) \dots \sum_{s_n} \exp\left(\lambda_n(0)(s_n - m_n)\right)\right\} \\
&= \ln \left\{(\exp(\lambda_i(0)) + 1) \exp(-\lambda_1(0) m_1) \dots (\exp(\lambda_i(0)) + 1) \exp(-\lambda_1(0) m_n)\right\} \\
&= \sum_i \left(\ln\left[1 + \left(\frac{m_i}{1 - m_i}\right)\right] - \lambda_i(0) m_i\right)
\end{aligned} \tag{14}$$

Using 13, we obtain:

$$\lambda_i(0) = \text{logit}(m_i) = \ln\left(\frac{m_i}{1 - m_i}\right)$$

and thus:

$$-(\beta G)_{\beta=0} = \sum_i \left\{ \ln\left(\frac{1}{1-m_i}\right) - m_i \ln\left(\frac{m_i}{1-m_i}\right) \right\}$$

$$= -\sum_i \left[m_i \ln(m_i) + (1-m_i) \ln(1-m_i) \right]$$

Yedida and Georges showed how to create the Taylor expansion beyond $O(\beta^2)$ (TODO referring to MF and TAP variant) (derivation in Appendix):

$$-\beta G = -\sum_i \left[m_i \ln(m_i) + (1-m_i) \ln(1-m_i) \right]$$

$$+ \beta \sum_{(ij)} w_{ij} m_i m_j + \beta \sum_i a_i m_i$$

$$+ \frac{\beta^2}{2} \sum_{(ij)} w_{ij}^2 (m_i - m_i^2)(m_j - m_j^2)$$

$$+ \frac{2\beta^3}{3} \sum_{(ij)} w_{ij}^3 (m_i - m_i^2)(\frac{1}{2} - m_i)(m_j - m_j^2)(\frac{1}{2} - m_j)$$

TODO:Add self consistency relations:

The variational free energy $\mathcal{F}$ is highly multimodal in real world examples in which case all stationary points either maximum, minimum or saddle satisfies the self-consistency relations **??**.

0.6. Marginal polytope - bounds for the partition function

Although it is very straightforward to obtain naive mean field approach from the TAP expansions, unlike the former this method doesn't bound in any way the variational free energy. Moreover, the approximation was based on the Taylor expansion which poses a threat that the radius of convergence of the expansion will be too small to obtain robust results for the different values of β [**yedidia2001idiosyncratic**]. There are a few examples in statistical physics where this method works very reliably in a wide variety of temperatures [**plefka1982convergence**] however in general there aren't any theoretical foundations for the robustness of the TAP expansion.

0.7. Boltzmann Machine

A Boltzmann machine is a network of symmetrically coupled stochastic binary units with both visible-to-visible and hidden-to-hidden lateral connections with energy function defined as: [**ackley1985learning**]

0.7.1. Approximator for any distribution

An example of such structure presents Figure **??**.

0.8. RBMs

Representational power - bengio -any distibution can be modelled by RBM. TODO.

$$E(\mathbf{v}, \mathbf{h}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{i,j} v_i w_{ij} h_j$$

$$P(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} e^{-E(\mathbf{v}, \mathbf{h})}$$

$$Z = \sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})}$$

$$F = -\ln Z$$

$$F^c(\mathbf{v}) = \ln\left(\sum_{\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h})}\right) = \ln\left(\sum_{h_1} e^{-E(\mathbf{v},h_1)}...\sum_{h_n} e^{-E(\mathbf{v},h_n)}\right)$$

$$\mathcal{L} = \ln P(\mathbf{v}) = F^c(\mathbf{v}) - F$$

where F is the *free energy* of the RBM.

0.8.1. Exploiting the RBM structure

The restrictions imposed on the structure allows for efficient computation of the clamped free energy because the hidden variables are independent given the state of the visible variables and vice versa and we can write:

$$p(\mathbf{h}|\mathbf{v}) = \prod_{i=1}^{m} p(h_i|\mathbf{v})$$
$$p(\mathbf{v}|\mathbf{h}) = \prod_{i=1}^{n} p(v_i|\mathbf{h})$$

(15)

Clamped free energy can be written in the form:

$$\begin{aligned}
\mathcal{F}^c(\mathbf{v}) &= \sum_{\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h})} = e^{\mathbf{b'v}} \sum_{h_1} ... \sum_{h_m} e^{-E(\mathbf{v},\mathbf{h})} \\
&= e^{\mathbf{b'v}} \sum_{h_1} e^{h_1(c_1 + W_1 \bullet \mathbf{v})} ... \sum_{h_m} e^{h_m(c_m + W_m \bullet \mathbf{v})} \\
&= e^{\mathbf{b'v}} \prod_{j=1}^{m} \left(1 + e^{c_i + W_i \bullet \mathbf{v}}\right)
\end{aligned}$$

(16)

Appendix

Following [**georges1991expand**] lets define energy as:

$$E = -\sum_{ij} w_{ij} s_i s_j - \sum_i a_i s_i$$

(17)

and introduce the following operator:

$$U \equiv E - \mathbb{E}(E) - \sum_i \frac{\partial \lambda_i(\beta)}{\partial \beta}(s_i - m_i)$$

(18)

which poses useful property:

$$\mathbb{E}(U) = 0$$

For any other operator O we then have:

$$\frac{\partial \mathbb{E}(O)}{\partial \beta} = \mathbb{E}\left(\frac{\partial O}{\partial \beta}\right) - \mathbb{E}(OU)$$

(19)

Now, the first derivative from the Taylor expansion is:

$$\begin{aligned}
\frac{\partial(\beta G)}{\partial \beta} &= \frac{\sum_{\mathbf{s}} \exp\left(\beta \sum_{(ij)} w_{ij} s_i s_j + \beta \sum_i a_i s_i + \sum_i \lambda_i(\beta)(s_i - m_i)\right)\left(\sum_{(ij)} w_{ij} s_i s_j + \sum_i a_i s_i + \sum_i \frac{\partial \lambda_i(\beta)}{\partial \beta}(s_i - m_i)\right)}{\sum_{\mathbf{s}} \exp\left(\beta \sum_{(ij)} w_{ij} s_i s_j + \sum_i a_i s_i + \sum_i \lambda_i(\beta)(s_i - m_i)\right)} \\
&= \sum_{(ij)} w_{ij} \mathbb{E}(s_i s_j) + \sum_i a_i \mathbb{E}(s_i) + \frac{\partial \lambda_i(\beta)}{\partial \beta} \sum_i \mathbb{E}(s_i - m_i)
\end{aligned}$$

In the case of $\beta = 0$ we have:
$$\left.\frac{\partial(\beta G)}{\partial \beta}\right|_{\beta=0} = \sum_{(ij)} w_{ij} m_i m_j + \sum_i a_i m_i$$

Using 19 we obtain: A.1.4:
$$\frac{\partial m_i}{\partial \beta} = 0 = \frac{\partial \mathbb{E}(s_i)}{\partial \beta} = \mathbb{E}\left(\frac{\partial s_i}{\partial \beta}\right) - \mathbb{E}(s_i U) = -\mathbb{E}(s_i U) = -\mathbb{E}(U(s_i - m_i)) \tag{20}$$

Thus A.1.5:
$$\begin{aligned}
\frac{\partial U}{\partial \beta} &= \frac{\partial H}{\partial \beta} - \frac{\partial \mathbb{E}(H)}{\partial \beta} - \sum_i \frac{\partial^2 \lambda_i(\beta)}{\partial \beta^2}(s_i - m_i) \\
&= \mathbb{E}(U^2) - \sum_i \frac{\partial^2 \lambda_i(\beta)}{\partial \beta^2}(s_i - m_i)
\end{aligned} \tag{21}$$

The second derivative has the form A.1.6:
$$\begin{aligned}
\frac{\partial^2 U}{\partial \beta^2} &= 2\mathbb{E}\left(\frac{\partial U}{\partial \beta} U\right) - \mathbb{E}(U^3) - \sum_i \frac{\partial^3 \lambda_i(\beta)}{\partial \beta^3}(s_i - m_i) \\
&= -\mathbb{E}(U^3) - \sum_i \frac{\partial^3 \lambda_i(\beta)}{\partial \beta^3}(s_i - m_i)
\end{aligned} \tag{22}$$

Coming back to our expansion of free energy using formulas derived above we now can calculate A.1.7:
$$\frac{\partial(\beta G)}{\partial \beta} = \mathbb{E}(E) - \sum_i \frac{\partial \lambda_i(\beta)}{\partial \beta} \mathbb{E}(s_i - m_i) = \mathbb{E}(E) \tag{23}$$

and A.1.8:
$$\frac{\partial^2(\beta G)}{\partial \beta^2} = \mathbb{E}\left(\frac{\partial E}{\partial \beta}\right) - \mathbb{E}(EU) = -\mathbb{E}(U^2)$$

$$\frac{\partial^3(\beta G)}{\partial \beta^3} = -2\mathbb{E}\left(U\frac{\partial U}{\partial \beta}\right) + \mathbb{E}(U^3) = \mathbb{E}(U^3) \tag{24}$$

$$\frac{\partial^4(\beta G)}{\partial \beta^4} = 3\mathbb{E}\left(U^2\frac{\partial U}{\partial \beta}\right) - \mathbb{E}(U^4) = 3\left(\mathbb{E}(U^2)\right)^2 - 3\sum_i \frac{\partial^2 \lambda_i(\beta)}{\partial \beta^2}\mathbb{E}(U^2(s_i - m_i)) - \mathbb{E}(U^4)$$

TS was considered around point $\beta = 0$. Using derivations from above we obtain again a 'naive' term:
$$\left.\frac{\partial(\beta G)}{\partial \beta}\right|_{\beta=0} = \mathbb{E}_{\beta=0}(E) = -\sum_{(ij)} w_{ij} m_i m_j - \sum_i a_i m_i = -\frac{1}{2}\sum_i \sum_j w_{ij} m_i m_j - \sum_i a_i m_i \tag{25}$$

Consider now
$$\left.\frac{\partial(\beta G)}{\partial m_i \partial \beta}\right|_{\beta=0} = -\sum_{j \neq i} w_{ij} m_j - a_i \tag{26}$$

On the other hand:
$$\left.\frac{\partial(\beta G)}{\partial m_i \partial \beta}\right|_{\beta=0} = \left.\frac{\partial(\beta G)}{\partial \beta \partial m_i}\right|_{\beta=0} = \frac{\partial}{\partial \beta}\mathbb{E}(\lambda_i(\beta)) = \left.\frac{\partial \lambda_i(\beta)}{\partial \beta}\right|_{\beta=0} \tag{27}$$

Substituting 27 into 18 gives us A.12:

$$\begin{aligned}
U_{\beta=0} &= -\sum_{(ij)} w_{ij} s_i s_j - \sum_i a_i s_i + \frac{1}{2}\sum_i \sum_j w_{ij} m_i m_j + \sum_i a_i m_i + \sum_i \left(\sum_{j \neq i} w_{ij} m_j + a_i\right)(s_i - m_i) \\
&= -\sum_{(ij)} w_{ij} s_i s_j - \frac{1}{2}\sum_{(ij)} w_{ij} m_i m_j + \sum_i \sum_{j \neq i} w_{ij} s_i m_j \\
&= -\sum_{(ij)} w_{ij}(s_i - m_i)(s_j - m_j) = -\sum_l w_l y_l
\end{aligned} \tag{28}$$

where y_l stands for the 'link' operator $w_l = w_{ij}$ and $y_l = (s_i - m_i)(s_j - m_j)$ which poses useful properties:

$$
\begin{aligned}
\mathbb{E}(y_l)_{\beta=0} &= \mathbb{E}(s_i s_j) - m_j \mathbb{E}(s_i) - m_i \mathbb{E}(s_j) + m_i m_j = 0 \\
\mathbb{E}(y_l (s_i - m_i))_{\beta=0} &= m_j - m_j - m_i^2 m_j + m_i^2 m_j \\
&\quad - m_i^2 m_j + m_i^2 m_j + m_i^2 m_j - m_i^2 m_j \\
&= 0
\end{aligned}
\tag{29}
$$

Finally, if $k \neq l$ then:
$$
\mathbb{E}(y_k y_l) = \mathbb{E}(y_k)\mathbb{E}(y_l) = 0
$$

while for $k = l$ we have:

$$
\begin{aligned}
\mathbb{E}((s_i - m_1)^2 (s_j - m_j)^2) &= m_i m_j - 2 m_i m_j^2 + m_i m_j^2 - 2 m_i^2 m_j + 4 m_1^2 m_j^2 \\
&\quad - 2 m_i^2 m_j^2 + m_i^2 m_j - 2 m_i^2 m_j^2 + m_i^2 m_j^2 \\
&= (m_i - m_i^2)(m_j - m_j^2)
\end{aligned}
\tag{30}
$$

Using properties from 30 in equations 24 we can derive **??** A1.13:

$$
\begin{aligned}
\left. \frac{\partial^2 (\beta G)}{\partial \beta^2} \right|_{\beta=0} &= -\mathbb{E}(U^2)_{\beta=0} \\
&= -\sum_{l_i l_2} w_{l_i} w_{l_2} \mathbb{E}_{\beta=0}(y_{l_1} y_{l_2}) \\
&= -\sum_{(i,j)} w_{ij}^2 (m_i - m_i^2)(m_j - m_j^2)
\end{aligned}
$$

which yields the TAP-Onsager term.

To obtain the next term for the Taylor expansion we need to compute $\mathbb{E}(y_{l_1} y_{l_2} y_{l_3})$ term and by definition the structure of the RBM model doesn't admit triangles in its corresponding factor graphs. Thus, we need to consider on the case when $l_1 = l_2 = l_3$:

$$
\begin{aligned}
\mathbb{E}((s_i - m_1)^3 (s_j - m_j)^3) &= m_i m_j - 3 m_i m_j^2 + 2 m_i m_j^3 + 2 m_i m_j^3 - 3 m_i^2 m_j + 2 m_i^3 m_j \\
&\quad + 9 m_i^2 m_j^2 - 6 m_i^3 m_j^2 - 6 m_i^2 m_j^3 + 4 m_i^3 m_j^3 \\
&= 4(m_i - m_i^2)(\tfrac{1}{2} - m_i)(m_j - m_j^2)(\tfrac{1}{2} - m_j)
\end{aligned}
\tag{31}
$$

$$
\left. \frac{\partial^3 (\beta G)}{\partial \beta^3} \right|_{\beta=0} = \frac{2\beta^3}{3} \sum_{(ij)} w_{ij}^3 (m_i - m_i^2)(\tfrac{1}{2} - m_i)(m_j - m_j^2)(\tfrac{1}{2} - m_j)
$$