Detailed derivation of Belief Propagation algorithm on Pairwise Markov Random Fields

Yurii Lahodiuk
yura.lagodiuk@gmail.com

1 Problem Statement

Lets consider a graph G – where cycles are absent (tree). One of the properties of G – is such, that it is a *bipartite* graph. It means, that there exists 2 disjoint sets of vertices (lets denote them as I and J), where edges between vertices of the same set are absent [1].

The set of edges of graph: E – can be represented as a set of tuples (i, j), where $i \in I$ and $j \in J$.

Also, lets introduce additional property of G: each vertex $i \in I$ is associated with some discrete valued random variable X_i, which has value $x_i \in S(X_i)$ (where, $S(X_i)$ is a finite set of possible outcomes of random variable X_i).

The same is applicable to all vertices from J – each vertex $j \in J$ is associated with discrete valued random variable X_j, which has value $x_j \in S(X_j)$.

Lets introduce shorthand notation for summation of some function $f(x_k)$, which defined $\forall x_k \in S(X_k)$ – over all outcomes of X_k:

$$(1) \qquad \sum_{x_k \in S(X_k)} f(x_k) \equiv \sum_{x_k} f(x_k), \qquad \forall k \in I \cup J$$

Lets assume that there exists some probability mass function, which represents probabilities of the possible "configurations" of graph [2]: $p(X_1 = x_1, X_2 = x_2, \dots X_n = x_n)$, where $x_1 \in S(X_1),\ x_2 \in S(X_2), \dots x_n \in S(X_n)$. Lets use following shorthand notation for expression of probability mass function:

$$(2) \qquad p(X_1 = x_1, X_2 = x_2, \dots X_n = x_n) \equiv p(x_1, x_2, \dots x_n)$$

By definition of probability mass function:

$$(3) \qquad \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} p(x_1, x_2, \dots x_n) = 1$$

We are interested in "single-node" marginal probabilities:

$$(4) \qquad \begin{aligned} p_i(x_i) &= \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} p(x_1, x_2, \dots x_n), \qquad \forall i \in I, \quad \forall x_i \in S(X_i) \\ p_j(x_i) &= \sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p(x_1, x_2, \dots x_n), \qquad \forall j \in J, \quad \forall x_j \in S(X_j) \end{aligned}$$

[1] Following intuition might be useful: imagine, that we assigned natural numbers: $1, 2, \dots n$ – for all vertices of G. So, we can state that I – represents set of vertices with *odd* indices, and J – represents set of vertices with *even* indices.

[2] Concept of "configuration" means: that random variable X_1, which associated with vertex 1, has some specific value x_1 – from the set of possible outcomes: $S(X_1)$; vertex 2 is associated with random variable X_2, which has some specific value x_2 – from the set of possible outcomes $S(X_2)$, and so on.

For nodes, which are connected – we, also interested in "pairwise" marginal probabilities:

$$
\begin{aligned}
&p_{ij}(x_i, x_j) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p(x_1, x_2, \dots x_n), \\
&\forall (i,j) \in E, \quad \forall x_i \in S(X_i), \quad \forall x_j \in S(X_j)
\end{aligned} \tag{5}
$$

Calculation of values of marginal probabilities, by straightforward summations of probability mass function over all possible outcomes of the random variables (except the fixed one) – leads to exponential complexity of computations:

$$
\begin{aligned}
&O\,(\ |S(X_1)| \times |S(X_2)| \times \cdots \times |S(X_{k-1})| \times |S(X_{k+1})| \times \cdots \times |S(X_n)|\ ) = O(\alpha^{n-1}) = O(\alpha^n), \\
&\forall k \in I \cup J, \\
&\alpha \text{ – is just some constant,} \\
&n \text{ – is number of vertices in graph } G
\end{aligned} \tag{6}
$$

In case if set of random variables (which associated with vertices of graph G) satisfies the local Markov Properties [1], and taking into account bipartiteness of G – we can state, that given set of random variables is, actually, a *Pairwise Markov Random Field.*

In the following section – we will expose *local properties* of the Pairwise Markov Random Field, which will help to reduce complexity of computations of the marginal probabilities.

2 Derivation

Keeping in mind Equation (3), lets consider following equalities for marginal probabilities [3] (which will be useful during the further steps of derivation):

$$
\begin{aligned}
\sum_{x_i} p_i(x_i) &= \sum_{x_i} \left(\sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} p(x_1, x_2, \dots x_n) \right) = \\
&= \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} p(x_1, x_2, \dots x_n) = 1 \\
\sum_{x_j} p_j(x_j) &= \sum_{x_j} \left(\sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p(x_1, x_2, \dots x_n) \right) = \\
&= \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} p(x_1, x_2, \dots x_n) = 1
\end{aligned} \tag{7}
$$

$$
\begin{aligned}
\sum_{x_i} \sum_{x_j} p_{ij}(x_i, x_j) &= \sum_{x_i} \sum_{x_j} \left(\sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p(x_1, x_2, \dots x_n) \right) = \\
&= \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} p(x_1, x_2, \dots x_n) = 1
\end{aligned} \tag{8}
$$

[3] Described equalities are possible, thanks to commutativity property of summations.

Lets represent single-node marginal probabilities, using pairwise marginal probabilities (keeping in mind Equation (5)):

$$
\begin{aligned}
p_i(x_i) &= \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} p(x_1, x_2, \ldots x_n) = \\
&= \sum_{x_j} \left(\sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p(x_1, x_2, \ldots x_n) \right) = \\
&= \sum_{x_j} p_{ij}(x_i, x_j) \\
p_j(x_j) &= \sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p(x_1, x_2, \ldots x_n) = \\
&= \sum_{x_i} \left(\sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p(x_1, x_2, \ldots x_n) \right) = \\
&= \sum_{x_i} p_{ij}(x_i, x_j)
\end{aligned} \tag{9}
$$

Lets introduce shorthand notations for summations over vertices (from I and J) and edges (from E):

$$
\begin{aligned}
\sum_{i \in I} f(i) &\equiv \sum_{i} f(i) \\
\sum_{j \in J} f(j) &\equiv \sum_{j} f(j) \\
\sum_{(i,j) \in E} f(i,j) &\equiv \sum_{(i,j)} f(i,j)
\end{aligned} \tag{10}
$$

Lets consider summation over all edges - of some functions $f_i(x_i)$ and $f_j(x_j)$ (which depends only on states of single nodes: i and j), also lets denote q_i – as a number of edges, which connected to node i ("degree" of node i), and q_j – "degree" of node j. So, in this case – we can represent summations over all edges, as summations over nodes, multiplied by node-degrees:

$$
\begin{aligned}
\sum_{(i,j)} f_i(x_i) &\equiv \sum_{i} \left(q_i \times f_i(x_i) \right), \quad \forall i \in I, \quad \forall x_i \in S(X_i) \\
\sum_{(i,j)} f_j(x_j) &\equiv \sum_{j} \left(q_j \times f_j(x_j) \right), \quad \forall j \in J, \quad \forall x_j \in S(X_j)
\end{aligned} \tag{11}
$$

Hammersley-Clifford Theorem [3, 4, 5] states, that probability distribution of configurations of Markov Random Field can be factorized into product of non-negative functions ("potentials") – defined over maximal cliques of graph:

$$
p(x_1, x_2, \ldots x_n) = \frac{1}{Z} \times \left(\prod_{(i,j)} \psi_{ij}(x_i, x_j) \right) \times \left(\prod_{i} \phi_i(x_i) \right) \times \left(\prod_{j} \phi_j(x_j) \right) \tag{12}
$$

Boltzmann's Law:

$$
p(x_1, x_2, \ldots x_n) = \frac{1}{Z} \times \exp\left(\frac{-E(x_1, x_2, \ldots x_n)}{T} \right)
$$

We can treat T just as a scaling coefficient, and for simplicity lets assume that $T = 1$:

$$p(x_1, x_2, \dots x_n) = \frac{1}{Z} \times \exp\left(-E(x_1, x_2, \dots x_n)\right) \tag{13}$$

From (12) and (13):

$$\exp\left(-E(x_1, x_2, \dots x_n)\right) = \left(\prod_{(i,j)} \psi_{ij}(x_i, x_j)\right) \times \left(\prod_{i} \phi_i(x_i)\right) \times \left(\prod_{j} \phi_j(x_j)\right)$$

So:

$$E(x_1, x_2, \dots x_n) = -\left(\sum_{(i,j)} ln(\psi_{ij}(x_i, x_j))\right) - \left(\sum_{i} ln(\phi_i(x_i))\right) - \left(\sum_{j} ln(\phi_j(x_j))\right) \tag{14}$$

Lets, assume that we have function $b(x_1, x_2, \dots x_n)$, which approximates real probability mass function $p(x_1, x_2, \dots x_n)$. It means, that Equations (3)–(5) and (7)–(9) the same for $b(x_1, x_2, \dots x_n)$:

$$\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) = 1 \tag{15}$$

$$\begin{aligned} b_i(x_i) &= \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \\ b_j(x_i) &= \sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \end{aligned} \tag{16}$$

$$b_{ij}(x_i, x_j) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \tag{17}$$

$$\begin{aligned} \sum_{x_i} b_i(x_i) &= 1 \\ \sum_{x_j} b_j(x_j) &= 1 \end{aligned} \tag{18}$$

$$\sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) = 1 \tag{19}$$

$$\begin{aligned} b_i(x_i) &= \sum_{x_j} b_{ij}(x_i, x_j) \\ b_j(x_j) &= \sum_{x_i} b_{ij}(x_i, x_j) \end{aligned} \tag{20}$$

Lets use Kullback-Leibler divergence for measurement of "difference" between real and approximated functions:

$$
\begin{aligned}
D_{KL}(b||p) &= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(b(x_1,x_2,\ldots x_n)\times ln\left(\frac{b(x_1,x_2,\ldots x_n)}{p(x_1,x_2,\ldots x_n)}\right)\right) = \\
&= \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\ldots x_n)\times ln\left(b(x_1,x_2,\ldots x_n)\right)\right) - \\
&- \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\ldots x_n)\times ln\left(p(x_1,x_2,\ldots x_n)\right)\right)
\end{aligned} \tag{21}
$$

Lest substitute $p(x_1,x_2,\ldots x_n)$ using expression from equation (13):

$$
\begin{aligned}
D_{KL}(b||p) &= \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\ldots x_n)\times ln\left(b(x_1,x_2,\ldots x_n)\right)\right) - \\
&- \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\ldots x_n)\times ln\left(\frac{1}{Z}\times\exp\left(-E(x_1,x_2,\ldots x_n)\right)\right)\right) = \\
&= \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\ldots x_n)\times ln\left(b(x_1,x_2,\ldots x_n)\right)\right) - \\
&- \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\ldots x_n)\times \left(-ln(Z)-E(x_1,x_2,\ldots x_n)\right)\right) = \\
&= \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\ldots x_n)\times ln\left(b(x_1,x_2,\ldots x_n)\right)\right) + \\
&+ \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\ldots x_n)\times ln(Z)\right) + \\
&+ \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\ldots x_n)\times E(x_1,x_2,\ldots x_n)\right)
\end{aligned} \tag{22}
$$

As far, as Z is just a constant – it doesn't depend on $x_1, x_2 \ldots x_n$, so we can move $ln(Z)$ out of summations:

$$
\begin{aligned}
D_{KL}(b||p) &= \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\ldots x_n)\times ln\left(b(x_1,x_2,\ldots x_n)\right)\right) + \\
&+ \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\ldots x_n)\right)\times ln(Z) + \\
&+ \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\ldots x_n)\times E(x_1,x_2,\ldots x_n)\right)
\end{aligned} \tag{23}
$$

Taking into account equation (15), we can substitute multiplier near $ln(Z)$ – to 1:

$$
\begin{aligned}
D_{KL}(b||p) &= \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\ldots x_n)\times ln\left(b(x_1,x_2,\ldots x_n)\right)\right) + \\
&+ \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\ldots x_n)\times E(x_1,x_2,\ldots x_n)\right) + ln(Z)
\end{aligned} \tag{24}
$$

Lets use following notation:

$$
\begin{aligned}
U(b(x_1, x_2, \dots x_n)) &= \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} (b(x_1, x_2, \dots x_n) \times E(x_1, x_2, \dots x_n)) \\
-H(b(x_1, x_2, \dots x_n)) &= \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} (b(x_1, x_2, \dots x_n) \times ln\,(b(x_1, x_2, \dots x_n))) \\
G(b(x_1, x_2, \dots x_n)) &= U(b(x_1, x_2, \dots x_n)) - H(b(x_1, x_2, \dots x_n)) \\
-F(b(x_1, x_2, \dots x_n)) &= -F = ln(Z)
\end{aligned}
\tag{25}
$$

So:

$$
D_{KL}(b||p) = G(b(x_1, x_2, \dots x_n)) - F = U(b(x_1, x_2, \dots x_n)) - H(b(x_1, x_2, \dots x_n)) + ln(Z) \tag{26}
$$

Let's transform $U(b(x_1, x_2, \dots x_n))$, using expression for energy (14):

(27)

$$U(b(x_1,x_2,\dots x_n)) = \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\Big(b(x_1,x_2,\dots x_n)\times E(x_1,x_2,\dots x_n)\Big) =$$

$$= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(b(x_1,x_2,\dots x_n)\times\left(-\left(\sum_{(i,j)} ln(\psi_{ij}(x_i,x_j))\right)-\left(\sum_{i} ln(\phi_i(x_i))\right)-\left(\sum_{j} ln(\phi_j(x_j))\right)\right)\right) =$$

$$= -\left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(b(x_1,x_2,\dots x_n)\times\sum_{(i,j)} ln(\psi_{ij}(x_i,x_j))\right)\right) -$$

$$-\left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(b(x_1,x_2,\dots x_n)\times\sum_{i} ln(\phi_i(x_i))\right)\right) -$$

$$-\left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(b(x_1,x_2,\dots x_n)\times\sum_{j} ln(\phi_j(x_j))\right)\right) =$$

$$= -\left(\sum_{(i,j)}\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\Big(b(x_1,x_2,\dots x_n)\times ln(\psi_{ij}(x_i,x_j))\Big)\right) -$$

$$-\left(\sum_{i}\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\Big(b(x_1,x_2,\dots x_n)\times ln(\phi_i(x_i))\Big)\right) -$$

$$-\left(\sum_{j}\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\Big(b(x_1,x_2,\dots x_n)\times ln(\phi_j(x_j))\Big)\right) =$$

$$= -\left(\sum_{(i,j)}\sum_{x_i}\sum_{x_j}\sum_{x_1}\cdots\sum_{x_{i-1}}\sum_{x_{i+1}}\cdots\sum_{x_{j-1}}\sum_{x_{j+1}}\cdots\sum_{x_n}(b(x_1,x_2,\dots x_n)\times ln(\psi_{ij}(x_i,x_j)))\right) -$$

$$-\left(\sum_{i}\sum_{x_i}\sum_{x_1}\cdots\sum_{x_{i-1}}\sum_{x_{i+1}}\cdots\sum_{x_n}(b(x_1,x_2,\dots x_n)\times ln(\phi_i(x_i)))\right) -$$

$$-\left(\sum_{j}\sum_{x_j}\sum_{x_1}\cdots\sum_{x_{j-1}}\sum_{x_{j+1}}\cdots\sum_{x_n}(b(x_1,x_2,\dots x_n)\times ln(\phi_j(x_j)))\right)$$

As far, as $\psi_{ij}(x_i,x_j)$ – depends only on x_i and x_j, and $\phi_i(x_i)$ – depends only on x_i, and $\phi_j(x_j)$ – depends only on x_j – we could rewrite sums as:

(28)

$$U(b(x_1,x_2,\dots x_n)) = -\left(\sum_{(i,j)}\sum_{x_i}\sum_{x_j} ln(\psi_{ij}(x_i,x_j)) \times \left(\sum_{x_1}\cdots\sum_{x_{i-1}}\sum_{x_{i+1}}\cdots\sum_{x_{j-1}}\sum_{x_{j+1}}\cdots\sum_{x_n} b(x_1,x_2,\dots x_n)\right)\right) -$$
$$-\left(\sum_{i}\sum_{x_i} ln(\phi_i(x_i)) \times \left(\sum_{x_1}\cdots\sum_{x_{i-1}}\sum_{x_{i+1}}\cdots\sum_{x_n} b(x_1,x_2,\dots x_n)\right)\right) -$$
$$-\left(\sum_{j}\sum_{x_j} ln(\phi_j(x_j)) \times \left(\sum_{x_1}\cdots\sum_{x_{j-1}}\sum_{x_{j+1}}\cdots\sum_{x_n} b(x_1,x_2,\dots x_n)\right)\right)$$

Taking into account Equations (16) and (17): we can substitute summations of probability mass function – to marginal probabilities:

$$\begin{aligned} U(b(x_1,x_2,\dots x_n)) &= -\left(\sum_{(i,j)}\sum_{x_i}\sum_{x_j} ln(\psi_{ij}(x_i,x_j)) \times b_{ij}(x_i,x_j)\right) - \\ &-\left(\sum_{i}\sum_{x_i} ln(\phi_i(x_i)) \times b_i(x_i)\right) - \left(\sum_{j}\sum_{x_j} ln(\phi_j(x_j)) \times b_j(x_j)\right) \end{aligned} \tag{29}$$

Lets introduce following variables ("local energies"):

$$\begin{aligned} &E_i(x_i) = -ln(\phi_i(x_i)) \\ &E_j(x_j) = -ln(\phi_j(x_j)) \\ &E_{ij}(x_i,x_j) = -ln(\psi_{ij}(x_i,x_j)) - ln(\phi_i(x_i)) - ln(\phi_j(x_j)) \end{aligned} \tag{30}$$

So, now we can express $ln(\phi_i(x_i))$, $ln(\phi_j(x_j))$ and $ln(\psi_{ij}(x_i,x_j))$ – using "local energies" (30):

$$\begin{aligned} &ln(\phi_i(x_i)) = -E_i(x_i) \\ &ln(\phi_j(x_j)) = -E_j(x_j) \\ &ln(\psi_{ij}(x_i,x_j)) = -E_{ij}(x_i,x_j) + E_i(x_i) + E_j(x_j) \end{aligned} \tag{31}$$

Lets transform $U(b(x_1,x_2,\dots x_n))$ using Equations (31):

$$\begin{aligned} U(b(x_1,x_2,\dots x_n)) &= -\left(\sum_{(i,j)}\sum_{x_i}\sum_{x_j} b_{ij}(x_i,x_j) \times (-E_{ij}(x_i,x_j) + E_i(x_i) + E_j(x_j))\right) - \\ &-\left(\sum_{i}\sum_{x_i} b_i(x_i) \times (-E_i(x_i))\right) - \left(\sum_{j}\sum_{x_j} b_j(x_j) \times (-E_j(x_j))\right) = \\ &= \left(\sum_{(i,j)}\sum_{x_i}\sum_{x_j} b_{ij}(x_i,x_j) \times E_{ij}(x_i,x_j)\right) - \left(\sum_{(i,j)}\sum_{x_i}\sum_{x_j} b_{ij}(x_i,x_j) \times E_i(x_i)\right) - \\ &-\left(\sum_{(i,j)}\sum_{x_j}\sum_{x_i} b_{ij}(x_i,x_j) \times E_j(x_j)\right) + \left(\sum_{i}\sum_{x_i} b_i(x_i) \times E_i(x_i)\right) + \left(\sum_{j}\sum_{x_j} b_j(x_j) \times E_j(x_j)\right) \end{aligned} \tag{32}$$

Taking into account that $E_i(x_i)$ is not depends on x_j, and $E_j(x_j)$ is not depends on x_i - we can move these multipliers out of corresponding summations:

$$
\begin{aligned}
U(b(x_1, x_2, \dots x_n)) = & \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times E_{ij}(x_i, x_j) \right) - \left(\sum_{(i,j)} \sum_{x_i} E_i(x_i) \times \sum_{x_j} b_{ij}(x_i, x_j) \right) - \\
& - \left(\sum_{(i,j)} \sum_{x_j} E_j(x_j) \times \sum_{x_i} b_{ij}(x_i, x_j) \right) + \left(\sum_{i} \sum_{x_i} b_i(x_i) \times E_i(x_i) \right) + \left(\sum_{j} \sum_{x_j} b_j(x_j) \times E_j(x_j) \right)
\end{aligned} \tag{33}
$$

Taking into account (20) – we can substitute summations over two-node marginal probabilities – to single-node marginal probabilities:

$$
\begin{aligned}
U(b(x_1, x_2, \dots x_n)) = & \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times E_{ij}(x_i, x_j) \right) - \left(\sum_{(i,j)} \sum_{x_i} E_i(x_i) \times b_i(x_i) \right) - \\
& - \left(\sum_{(i,j)} \sum_{x_j} E_j(x_j) \times b_j(x_j) \right) + \left(\sum_{i} \sum_{x_i} b_i(x_i) \times E_i(x_i) \right) + \left(\sum_{j} \sum_{x_j} b_j(x_j) \times E_j(x_j) \right)
\end{aligned} \tag{34}
$$

Finally, we can substitute summations over edges – to summation over nodes, as described by Equation (11):

$$
\begin{aligned}
U(b(x_1, x_2, \dots x_n)) = & \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times E_{ij}(x_i, x_j) \right) - \left(\sum_{x_i} \sum_{(i,j)} E_i(x_i) \times b_i(x_i) \right) - \\
& - \left(\sum_{x_j} \sum_{(i,j)} E_j(x_j) \times b_j(x_j) \right) + \left(\sum_{i} \sum_{x_i} b_i(x_i) \times E_i(x_i) \right) + \left(\sum_{j} \sum_{x_j} b_j(x_j) \times E_j(x_j) \right) = \\
& = \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times E_{ij}(x_i, x_j) \right) - \left(\sum_{x_i} \sum_{i} q_i \times E_i(x_i) \times b_i(x_i) \right) - \\
& - \left(\sum_{x_j} \sum_{j} q_j \times E_j(x_j) \times b_j(x_j) \right) + \left(\sum_{i} \sum_{x_i} b_i(x_i) \times E_i(x_i) \right) + \left(\sum_{j} \sum_{x_j} b_j(x_j) \times E_j(x_j) \right) = \\
& = \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times E_{ij}(x_i, x_j) \right) - \\
& - \left(\sum_{i} \sum_{x_i} (q_i - 1) \times E_i(x_i) \times b_i(x_i) \right) - \left(\sum_{j} \sum_{x_j} (q_j - 1) \times E_j(x_j) \times b_j(x_j) \right)
\end{aligned} \tag{35}
$$

So, finally, we derived representation of $U(b(x_1, x_2, \dots x_n))$ in terms of marginal probabilities and "local energies":

$$
\begin{aligned}
U(b(x_1,x_2,\dots x_n)) &= \left(\sum_{(i,j)}\sum_{x_i}\sum_{x_j} b_{ij}(x_i,x_j)\times E_{ij}(x_i,x_j)\right) - \\
&- \left(\sum_{i}\sum_{x_i}(q_i-1)\times E_i(x_i)\times b_i(x_i)\right) - \left(\sum_{j}\sum_{x_j}(q_j-1)\times E_j(x_j)\times b_j(x_j)\right)
\end{aligned} \tag{36}
$$

Now, lets transform $-H(b(x_1,x_2,\dots x_n))$ from Equations (25):

$$
-H(b(x_1,x_2,\dots x_n)) = \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(b(x_1,x_2,\dots x_n)\times ln\left(b(x_1,x_2,\dots x_n)\right)\right) \tag{37}
$$

Lets introduce additional statement [6] (which can be proven by induction [7]) – probability distribution can be expressed via marginal probabilities (over max-cliques and single nodes), and node-degrees of graph. For Pairwise Markov Random Field this statement can be formalised as:

$$
b(x_1,x_2,\dots x_n) = \frac{\prod_{(i,j)} b_{ij}(x_i,x_j)}{\prod_i b_i(x_i)^{q_i-1}\times\prod_j b_j(x_j)^{q_j-1}} \tag{38}
$$

So, lets rewrite Equation (37), keeping in mind Equation (38):

$$
\begin{aligned}
&-H(b(x_1,x_2,\dots x_n)) = \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(b(x_1,x_2,\dots x_n)\times ln\left(b(x_1,x_2,\dots x_n)\right)\right) = \\
&= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(b(x_1,x_2,\dots x_n)\times ln\left(\frac{\prod_{(i,j)} b_{ij}(x_i,x_j)}{\prod_i b_i(x_i)^{q_i-1}\times\prod_j b_j(x_j)^{q_j-1}}\right)\right) = \\
&= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(b(x_1,x_2,\dots x_n)\times\left(ln\left(\prod_{(i,j)} b_{ij}(x_i,x_j)\right) - ln\left(\prod_i b_i(x_i)^{q_i-1}\right) - ln\left(\prod_j b_j(x_j)^{q_j-1}\right)\right)\right) = \\
&= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(b(x_1,x_2,\dots x_n)\times\left(\left(\sum_{(i,j)} ln\left(b_{ij}(x_i,x_j)\right)\right) - \left(\sum_i ln\left(b_i(x_i)^{q_i-1}\right)\right) - \left(\sum_j ln\left(b_j(x_j)^{q_j-1}\right)\right)\right)\right) = \\
&= \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\dots x_n)\times\sum_{(i,j)} ln\left(b_{ij}(x_i,x_j)\right)\right) - \\
&- \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\dots x_n)\times\sum_i ln\left(b_i(x_i)^{q_i-1}\right)\right) - \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\dots x_n)\times\sum_j ln\left(b_j(x_j)^{q_j-1}\right)\right) = \\
&= \left(\sum_{(i,j)}\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\dots x_n)\times ln\left(b_{ij}(x_i,x_j)\right)\right) - \\
&- \left(\sum_i\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\dots x_n)\times ln\left(b_i(x_i)^{q_i-1}\right)\right) - \left(\sum_j\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1,x_2,\dots x_n)\times ln\left(b_j(x_j)^{q_j-1}\right)\right)
\end{aligned} \tag{39}
$$

Lets rearrange sums, according to dependance of marginal probabilities:

(40)

$$
\begin{aligned}
-H(b(x_1, x_2, \dots x_n)) = & \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \times ln\left(b_{ij}(x_i, x_j)\right) \right) - \\
& - \left(\sum_{i} \sum_{x_i} \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \times ln\left(b_i(x_i)^{q_i-1}\right) \right) - \\
& - \left(\sum_{j} \sum_{x_j} \sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \times ln\left(b_j(x_j)^{q_j-1}\right) \right) = \\
& = \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} ln\left(b_{ij}(x_i, x_j)\right) \times \left(\sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \right) \right) - \\
& - \left(\sum_{i} \sum_{x_i} ln\left(b_i(x_i)^{q_i-1}\right) \times \left(\sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \right) \right) - \\
& - \left(\sum_{j} \sum_{x_j} ln\left(b_j(x_j)^{q_j-1}\right) \times \left(\sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \right) \right)
\end{aligned}
$$

Again, taking into account Equations (16) and (17): we can substitute summations of probability mass function – to marginal probabilities:

(41)

$$
\begin{aligned}
-H(b(x_1, x_2, \dots x_n)) = & \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} ln\left(b_{ij}(x_i, x_j)\right) \times \left(\sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \right) \right) - \\
& - \left(\sum_{i} \sum_{x_i} ln\left(b_i(x_i)^{q_i-1}\right) \times \left(\sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \right) \right) - \\
& - \left(\sum_{j} \sum_{x_j} ln\left(b_j(x_j)^{q_j-1}\right) \times \left(\sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \right) \right) = \\
& = \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} ln\left(b_{ij}(x_i, x_j)\right) \times b_{ij}(x_i, x_j) \right) - \\
& - \left(\sum_{i} \sum_{x_i} ln\left(b_i(x_i)^{q_i-1}\right) \times b_i(x_i) \right) - \left(\sum_{j} \sum_{x_j} ln\left(b_j(x_j)^{q_j-1}\right) \times b_j(x_j) \right)
\end{aligned}
$$

So, finally, we can represent $-H(b(x_1, x_2, \dots x_n))$ only in terms of marginal probabilities:

(42)
$$-H(b(x_1, x_2, \dots x_n)) = \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} ln\left(b_{ij}(x_i, x_j)\right) \times b_{ij}(x_i, x_j)\right) -$$
$$-\left(\sum_{i} \sum_{x_i} (q_i - 1) \times ln(b_i(x_i)) \times b_i(x_i)\right) - \left(\sum_{j} \sum_{x_j} (q_j - 1) \times ln(b_j(x_j)) \times b_j(x_j)\right)$$

Lets represent Kullback-Leibler divergence (26) – in terms of marginal probabilities and local energies (keeping in mind Equations (36) and (42)):

(43)
$$D_{KL}(b||p) = U(b(x_1, x_2, \dots x_n)) - H(b(x_1, x_2, \dots x_n)) + ln(Z) =$$
$$= \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times E_{ij}(x_i, x_j)\right) -$$
$$-\left(\sum_{i} \sum_{x_i} (q_i - 1) \times E_i(x_i) \times b_i(x_i)\right) - \left(\sum_{j} \sum_{x_j} (q_j - 1) \times E_j(x_j) \times b_j(x_j)\right) +$$
$$+ \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} ln\left(b_{ij}(x_i, x_j)\right) \times b_{ij}(x_i, x_j)\right) -$$
$$-\left(\sum_{i} \sum_{x_i} (q_i - 1) \times ln(b_i(x_i)) \times b_i(x_i)\right) - \left(\sum_{j} \sum_{x_j} (q_j - 1) \times ln(b_j(x_j)) \times b_j(x_j)\right) + ln(Z) =$$
$$= \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times \left(E_{ij}(x_i, x_j) + ln\left(b_{ij}(x_i, x_j)\right)\right)\right) -$$
$$-\left(\sum_{i} \sum_{x_i} (q_i - 1) \times b_i(x_i) \times (E_i(x_i) + ln(b_i(x_i)))\right) - \left(\sum_{j} \sum_{x_j} (q_j - 1) \times b_j(x_j) \times (E_j(x_j) + ln(b_j(x_j)))\right) + ln(Z)$$

So, again, I would like to pay attention on the fact – that we avoided potentially exponential amount of operations, during calculation of Kullback-Leibler divergence:

(44)
$$D_{KL}(b||p) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} \left(b(x_1, x_2, \dots x_n) \times ln\left(\frac{b(x_1, x_2, \dots x_n)}{p(x_1, x_2, \dots x_n)}\right)\right) =$$
$$= \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times \left(E_{ij}(x_i, x_j) + ln\left(b_{ij}(x_i, x_j)\right)\right)\right) -$$
$$-\left(\sum_{i} \sum_{x_i} (q_i - 1) \times b_i(x_i) \times (E_i(x_i) + ln(b_i(x_i)))\right) - \left(\sum_{j} \sum_{x_j} (q_j - 1) \times b_j(x_j) \times (E_j(x_j) + ln(b_j(x_j)))\right) + ln(Z)$$

So, our goal is to find such values of $b_i(x_i)$, $b_j(x_j)$ and $b_{ij}(x_i, x_j)$ – which leads to minimal value of $D_{KL}(b||p)$, with respect to conditions, defined by Equations (18)–(20).

Lets use Lagrange multipliers method [9] for this purpose. Lets construct lagrangian, according to restrictions from Equations (18)–(20):

$$
\begin{aligned}
\mathcal{L} = D_{KL}(b||p) &+ \left(\sum_i \gamma_i \times \left(1 - \sum_{x_i} b_i(x_i)\right)\right) + \left(\sum_j \gamma_j \times \left(1 - \sum_{x_j} b_j(x_j)\right)\right) + \\
&+ \left(\sum_{(i,j)} \gamma_{ij} \times \left(1 - \sum_{x_i}\sum_{x_j} b_{ij}(x_i, x_j)\right)\right) + \left(\sum_i \sum_{x_i} \sum_{j \in N(i)} \lambda_{ji}(x_i) \times \left(b_i(x_i) - \sum_{x_j} b_{ij}(x_i, x_j)\right)\right) + \\
&+ \left(\sum_j \sum_{x_j} \sum_{i \in N(j)} \lambda_{ij}(x_j) \times \left(b_j(x_j) - \sum_{x_i} b_{ij}(x_i, x_j)\right)\right)
\end{aligned} \tag{45}
$$

Where:

- γ_i, γ_j, γ_{ij}, $\lambda_{ji}(x_i)$ and $\lambda_{ij}(x_j)$ – lagrange multipliers (just some constants)
- $\sum_{j \in N(i)}$ – sum over node-indices j, which are adjacent to node i ("neighbours")
  Which means, that $\forall j \in N(i)$ exists edge (which represented by tuple (i, j))
- $\sum_{i \in N(j)}$ – sum over node-indices i, which are adjacent to node j

So, $\mathcal{L}$ – is a function, which depends on variables:

- $b_i(x_i)$ – defined $\forall i \in OddNodes$, and $\forall x_i \in StatesOfNode(i)$
- $b_j(x_j)$ – defined $\forall j \in EvenNodes$, and $\forall x_j \in StatesOfNode(j)$
- $b_{ij}(x_i, x_j)$ – defined $\forall (i, j) \in Edges$, and $\forall x_i \in StatesOfNode(i)$, and $\forall x_j \in StatesOfNode(j)$

By definition of Lagrange multipliers method, we are interested in stationary points of $\mathcal{L}$:

$$
\frac{\partial \mathcal{L}}{\partial b_{ij}(x_i, x_j)} = 0, \ (\forall (i, j) \in Edges, \ \forall x_i \in StatesOfNode(i), \ \forall x_j \in StatesOfNode(j)) \tag{46}
$$

$$
\frac{\partial \mathcal{L}}{\partial b_i(x_i)} = 0, \ (\forall i \in OddNodes, \ \forall x_i \in StatesOfNode(i)) \tag{47}
$$

$$
\frac{\partial \mathcal{L}}{\partial b_j(x_j)} = 0, \ (\forall j \in EvenNodes, \ \forall x_j \in StatesOfNode(j)) \tag{48}
$$

Lets transform Equation (46), using Equation (45):

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial b_{ij}(x_i, x_j)} &= \frac{\partial \left(D_{KL}(b||p)\right)}{\partial b_{ij}(x_i, x_j)} + \frac{\partial \left(- \gamma_{ij} \times b_{ij}(x_i, x_j)\right)}{\partial b_{ij}(x_i, x_j)} + \\
&+ \frac{\partial \left(- \lambda_{ji}(x_i) \times b_{ij}(x_i, x_j)\right)}{\partial b_{ij}(x_i, x_j)} + \frac{\partial \left(- \lambda_{ij}(x_j) \times b_{ij}(x_i, x_j)\right)}{\partial b_{ij}(x_i, x_j)} = \\
&= \frac{\partial \left(D_{KL}(b||p)\right)}{\partial b_{ij}(x_i, x_j)} - \gamma_{ij} - \lambda_{ji}(x_i) - \lambda_{ij}(x_j) = 0
\end{aligned} \tag{49}
$$

For the further transformation of partial derivative – lets use expression for $D_{KL}(b||p)$ from Equation (44):

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial b_{ij}(x_i,x_j)} &= \frac{\partial\left(D_{KL}(b||p)\right)}{\partial b_{ij}(x_i,x_j)} - \gamma_{ij} - \lambda_{ji}(x_i) - \lambda_{ij}(x_j) = \\
&= \frac{\partial\left(b_{ij}(x_i,x_j) \times \left(E_{ij}(x_i,x_j) + ln\left(b_{ij}(x_i,x_j)\right)\right)\right)}{\partial b_{ij}(x_i,x_j)} - \gamma_{ij} - \lambda_{ji}(x_i) - \lambda_{ij}(x_j) = \\
&= E_{ij}(x_i,x_j) + \frac{\partial\left(b_{ij}(x_i,x_j) \times ln\left(b_{ij}(x_i,x_j)\right)\right)}{\partial b_{ij}(x_i,x_j)} - \gamma_{ij} - \lambda_{ji}(x_i) - \lambda_{ij}(x_j) = \\
&= E_{ij}(x_i,x_j) + ln\left(b_{ij}(x_i,x_j)\right) + 1 - \gamma_{ij} - \lambda_{ji}(x_i) - \lambda_{ij}(x_j) = 0
\end{aligned} \tag{50}
$$

So, finally, from Equation (50) – we can derive expression for $ln\left(b_{ij}(x_i,x_j)\right)$:

$$ln\left(b_{ij}(x_i,x_j)\right) = \gamma_{ij} + \lambda_{ji}(x_i) + \lambda_{ij}(x_j) - E_{ij}(x_i,x_j) - 1 \tag{51}$$

Now, lets transform Equation (47), using Equation (45):

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial b_i(x_i)} &= \frac{\partial\left(D_{KL}(b||p)\right)}{\partial b_i(x_i)} + \frac{\partial\left(-\gamma_i \times b_i(x_i)\right)}{\partial b_i(x_i)} + \frac{\partial\left(\sum_{j \in N(i)} \lambda_{ji}(x_i) \times b_i(x_i)\right)}{\partial b_i(x_i)} = \\
&= \frac{\partial\left(D_{KL}(b||p)\right)}{\partial b_i(x_i)} - \gamma_i + \sum_{j \in N(i)} \left(\frac{\partial\left(\lambda_{ji}(x_i) \times b_i(x_i)\right)}{\partial b_i(x_i)}\right) = \\
&= \frac{\partial\left(D_{KL}(b||p)\right)}{\partial b_i(x_i)} - \gamma_i + \sum_{j \in N(i)} \lambda_{ji}(x_i) = 0
\end{aligned} \tag{52}
$$

Again, for the further transformation of partial derivative – lets use expression for $D_{KL}(b||p)$ from Equation (44):

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial b_i(x_i)} &= \frac{\partial\left(-(q_i-1) \times b_i(x_i) \times \left(E_i(x_i) + ln(b_i(x_i))\right)\right)}{\partial b_i(x_i)} - \gamma_i + \sum_{j \in N(i)} \lambda_{ji}(x_i) = \\
&= -(q_i-1) \times E_i(x_i) - (q_i-1) \times \frac{\partial\left(b_i(x_i) \times ln(b_i(x_i))\right)}{\partial b_i(x_i)} - \gamma_i + \sum_{j \in N(i)} \lambda_{ji}(x_i) = \\
&= -(q_i-1) \times E_i(x_i) - (q_i-1) \times \left(1 + ln(b_i(x_i))\right) - \gamma_i + \sum_{j \in N(i)} \lambda_{ji}(x_i) = 0
\end{aligned} \tag{53}
$$

So, finally, from Equation (53) – we can derive expression for $ln\left(b_i(x_i)\right)$:

$$(q_i - 1) \times (1 + ln(b_i(x_i))) = -(q_i - 1) \times E_i(x_i) - \gamma_i + \sum_{j \in N(i)} \lambda_{ji}(x_i)$$

$$(1 + ln(b_i(x_i))) = -E_i(x_i) - \frac{\gamma_i}{q_i - 1} + \frac{\sum_{j \in N(i)} \lambda_{ji}(x_i)}{q_i - 1} \tag{54}$$

$$ln(b_i(x_i)) = -E_i(x_i) - \frac{\gamma_i}{q_i - 1} - 1 + \frac{1}{q_i - 1} \times \sum_{j \in N(i)} \lambda_{ji}(x_i)$$

In the identical way – we can derive expression for $ln\left(b_j(x_j)\right)$:

$$ln(b_j(x_j)) = -E_j(x_j) - \frac{\gamma_j}{q_j - 1} - 1 + \frac{1}{q_j - 1} \times \sum_{i \in N(j)} \lambda_{ij}(x_j) \tag{55}$$

So, finally, we derived expressions for logarithms of marginal probabilities, which corresponds to stationary points of lagrangian:

$$\begin{cases} ln\left(b_{ij}(x_i, x_j)\right) = \gamma_{ij} + \lambda_{ji}(x_i) + \lambda_{ij}(x_j) - E_{ij}(x_i, x_j) - 1 \\ ln(b_i(x_i)) = -E_i(x_i) - \frac{\gamma_i}{q_i - 1} - 1 + \frac{1}{q_i - 1} \times \sum_{j \in N(i)} \lambda_{ji}(x_i) \\ ln(b_j(x_j)) = -E_j(x_j) - \frac{\gamma_j}{q_j - 1} - 1 + \frac{1}{q_j - 1} \times \sum_{i \in N(j)} \lambda_{ij}(x_j) \end{cases} \tag{56}$$

As you remember (from Equation (45)): γ_i, γ_j, γ_{ij}, $\lambda_{ji}(x_i)$ and $\lambda_{ij}(x_j)$ – are just some constants. So, if we find a way to compute these constants we will be able to compute marginal probabilities.

Lets represent $\lambda_{ji}(x_i)$ – as a sum of $(q_i - 1)$ constants:

$$\lambda_{ji}(x_i) = \sum_{k \in N(i) \setminus j} ln\left(m_{k \to i}(x_i)\right) \tag{57}$$

As far as node i has q_i neighbours, we can imagine, that each constant $ln\left(m_{k \to i}(x_i)\right)$ – belongs to edge (k, i) (except edge (j, i)).

The same is applied to $\lambda_{ij}(x_j)$:

$$\lambda_{ij}(x_j) = \sum_{k \in N(j) \setminus i} ln\left(m_{k \to j}(x_j)\right) \tag{58}$$

Lets rewrite Equations (56), using Equations (57) and (58):

$$
(59)\quad \begin{cases}
ln\left(b_{ij}(x_i,x_j)\right) = -E_{ij}(x_i,x_j) + \gamma_{ij} - 1 + \left(\sum\limits_{k\in N(i)\backslash j} ln\left(m_{k\to i}(x_i)\right)\right) + \left(\sum\limits_{k\in N(j)\backslash i} ln\left(m_{k\to j}(x_j)\right)\right) \\
ln(b_i(x_i)) = -E_i(x_i) - \dfrac{\gamma_i}{q_i - 1} - 1 + \dfrac{1}{q_i-1} \times \sum\limits_{j\in N(i)} \left(\sum\limits_{k\in N(i)\backslash j} ln\left(m_{k\to i}(x_i)\right)\right) \\
ln(b_j(x_j)) = -E_j(x_j) - \dfrac{\gamma_j}{q_j - 1} - 1 + \dfrac{1}{q_j-1} \times \sum\limits_{i\in N(j)} \left(\sum\limits_{k\in N(j)\backslash i} ln\left(m_{k\to j}(x_j)\right)\right)
\end{cases}
$$

Also, it can be easy shown that:

$$
(60)\quad \begin{aligned}
&\frac{1}{q_i-1} \times \sum_{j\in N(i)} \sum_{k\in N(i)\backslash j} f(k) \equiv \sum_{k\in N(i)} f(k) \\
&\frac{1}{q_j-1} \times \sum_{i\in N(j)} \sum_{k\in N(j)\backslash i} f(k) \equiv \sum_{k\in N(j)} f(k)
\end{aligned}
$$

So, finally we can represent logarithms of marginal probabilities as:

$$
(61)\quad \begin{cases}
ln\left(b_{ij}(x_i,x_j)\right) = -E_{ij}(x_i,x_j) + \gamma_{ij} - 1 + \left(\sum\limits_{k\in N(i)\backslash j} ln\left(m_{k\to i}(x_i)\right)\right) + \left(\sum\limits_{k\in N(j)\backslash i} ln\left(m_{k\to j}(x_j)\right)\right) \\
ln(b_i(x_i)) = -E_i(x_i) - \dfrac{\gamma_i}{q_i - 1} - 1 + \sum\limits_{k\in N(i)} ln\left(m_{k\to i}(x_i)\right) \\
ln(b_j(x_j)) = -E_j(x_j) - \dfrac{\gamma_j}{q_j - 1} - 1 + \sum\limits_{k\in N(j)} ln\left(m_{k\to j}(x_j)\right)
\end{cases}
$$

Lets substitute local energies ($E_i(x_i)$, $E_j(x_j)$ and $E_{ij}(x_i,x_j)$) – using Equations (30):

$$
(62)\quad \begin{cases}
ln\left(b_{ij}(x_i,x_j)\right) = ln(\psi_{ij}(x_i,x_j)) + ln(\phi_i(x_i)) + ln(\phi_j(x_j)) + \gamma_{ij} - 1 + \\
\qquad + \left(\sum\limits_{k\in N(i)\backslash j} ln\left(m_{k\to i}(x_i)\right)\right) + \left(\sum\limits_{k\in N(j)\backslash i} ln\left(m_{k\to j}(x_j)\right)\right) \\
ln(b_i(x_i)) = ln(\phi_i(x_i)) - \dfrac{\gamma_i}{q_i - 1} - 1 + \sum\limits_{k\in N(i)} ln\left(m_{k\to i}(x_i)\right) \\
ln(b_j(x_j)) = ln(\phi_j(x_j)) - \dfrac{\gamma_j}{q_j - 1} - 1 + \sum\limits_{k\in N(j)} ln\left(m_{k\to j}(x_j)\right)
\end{cases}
$$

Now, lets exponentiate Equations (62) – to get expressions for marginal probabilities:

$$
(63)\quad \begin{cases}
b_{ij}(x_i,x_j) = \exp(\gamma_{ij}-1) \times \left(\prod\limits_{k\in N(i)\setminus j} m_{k\to i}(x_i)\right) \times \phi_i(x_i) \times \psi_{ij}(x_i,x_j) \times \phi_j(x_j) \times \left(\prod\limits_{k\in N(j)\setminus i} m_{k\to j}(x_j)\right) \\
b_i(x_i) = \exp\left(-\dfrac{\gamma_i}{q_i-1}-1\right) \times \phi_i(x_i) \times \left(\prod\limits_{k\in N(i)} m_{k\to i}(x_i)\right) \\
b_j(x_j) = \exp\left(-\dfrac{\gamma_j}{q_j-1}-1\right) \times \phi_j(x_j) \times \left(\prod\limits_{k\in N(j)} m_{k\to j}(x_j)\right)
\end{cases}
$$

If we look at Equations (63) – we can notice, that following multipliers does not depend on states of graph nodes (x_i and x_j):

$$
(64)\quad \begin{cases}
\exp(\gamma_{ij}-1) \neq f(x_i,x_j) \\
\exp\left(-\dfrac{\gamma_i}{q_i-1}-1\right) \neq f(x_i) \\
\exp\left(-\dfrac{\gamma_j}{q_j-1}-1\right) \neq f(x_j)
\end{cases}
$$

Lets denote:

$$
(65)\quad \begin{cases}
Z_{ij} = \exp(\gamma_{ij}-1) \\
Z_i = \exp\left(-\dfrac{\gamma_i}{q_i-1}-1\right) \\
Z_j = \exp\left(-\dfrac{\gamma_j}{q_j-1}-1\right)
\end{cases}
$$

So, actually, it means, that given multipliers just corresponds to normalization coefficients (with respect to Equations (18) and (19)):

$$
(66)\quad \begin{cases}
b_{ij}(x_i,x_j) = Z_{ij} \times \left(\prod\limits_{k\in N(i)\setminus j} m_{k\to i}(x_i)\right) \times \phi_i(x_i) \times \psi_{ij}(x_i,x_j) \times \phi_j(x_j) \times \left(\prod\limits_{k\in N(j)\setminus i} m_{k\to j}(x_j)\right), \\
\qquad \text{where } Z_{ij} \text{ is such that } \sum\limits_{x_i}\sum\limits_{x_j} b_{ij}(x_i,x_j) = 1 \\
b_i(x_i) = Z_i \times \phi_i(x_i) \times \left(\prod\limits_{k\in N(i)} m_{k\to i}(x_i)\right), \\
\qquad \text{where } Z_i \text{ is such that } \sum\limits_{x_i} b_i(x_i) = 1 \\
b_j(x_j) = Z_j \times \phi_j(x_j) \times \left(\prod\limits_{k\in N(j)} m_{k\to j}(x_j)\right), \\
\qquad \text{where } Z_j \text{ is such that } \sum\limits_{x_j} b_j(x_j) = 1
\end{cases}
$$

So, if we find a way to calculate appropriate values of $m_{k\to i}(x_i)$ ($\forall i \in I,\ \ \forall k \in N(i),\ \ \forall x_i \in S(X_i)$) and $m_{k\to j}(x_j)$ ($\forall j \in J,\ \ \forall k \in N(j),\ \ \forall x_j \in S(X_j)$) – we will be able to calculate marginal probabilities $b_i(x_i)$ ($\forall i \in I,\ \ \forall x_i \in S(X_i)$) and $b_j(x_j)$ ($\forall j \in J,\ \ \forall x_j \in S(X_j)$).

References

[1] http://en.wikipedia.org/wiki/Markov_random_field

[2] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss, *Understanding Belief Propagation and its Generalizations*, Mitsubishi Electric Research Laboratories, TR2001-22, November 2001, http://www.merl.com/publications/docs/TR2001-22.pdf

[3] J. M. Hammersley, P. Clifford, *Markov fields on finite graphs and lattices*, 1971, http://www.statslab.cam.ac.uk/~grg/books/hammfest/hamm-cliff.pdf.

[4] Samson Cheung, *Proof of Hammersley-Clifford Theorem*, February 3, 2008, http://web.kaist.ac.kr/~kyomin/Fall09MRF/Hammersley-Clifford_Theorem.pdf

[5] http://en.wikipedia.org/wiki/Hammersley-Clifford_theorem

[6] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss, *Bethe free energy, Kikuchi approximations, and belief propagation algorithms*, Mitsubishi Electric Research Laboratories, TR-2001-16, May 2001, http://www.merl.com/publications/docs/TR2001-16.pdf

[7] Andrea Montanari, *Lecture Notes for Stat 375 Inference in Graphical Models*, Department of Electrical Engineering and Department of Statistics, Stanford University, February 21, 2011, (page 41), http://web.stanford.edu/~montanar/TEACHING/Stat375/handouts/notes_stat375_1.pdf

[8] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss, *Constructing Free Energy Approximations and Generalized Belief Propagation Algorithms*, Mitsubishi Electric Research Laboratories, TR2004-040, May 2004, http://www.merl.com/publications/docs/TR2004-040.pdf

[9] http://en.wikipedia.org/wiki/Lagrange_multiplier

[10] http://en.wikipedia.org/wiki/Fixed-point_iteration