Detailed derivation of Belief Propagation algorithm on Pairwise Markov Random Fields

Yurii Lahodiuk

yura.lagodiuk@gmail.com

By definition of probability density function:

$$(1) \quad \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} p(x_1, x_2, \ldots x_n) = 1$$

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

$$(2)$$
$$p_i(x_i) = \sum_{x_1}\cdots\sum_{x_{i-1}}\sum_{x_{i+1}}\cdots\sum_{x_n} p(x_1, x_2, \ldots x_n)$$

$$p_j(x_i) = \sum_{x_1}\cdots\sum_{x_{j-1}}\sum_{x_{j+1}}\cdots\sum_{x_n} p(x_1, x_2, \ldots x_n)$$

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

$$(3) \quad p_{ij}(x_i, x_j) = \sum_{x_1}\cdots\sum_{x_{i-1}}\sum_{x_{i+1}}\cdots\sum_{x_{j-1}}\sum_{x_{j+1}}\cdots\sum_{x_n} p(x_1, x_2, \ldots x_n)$$

Also, the following equalities for marginal probabilities – are obvious (keeping in mind Equation (1)):

$$(4)$$
$$\sum_{x_i} p_i(x_i) = \sum_{x_i}\left(\sum_{x_1}\cdots\sum_{x_{i-1}}\sum_{x_{i+1}}\cdots\sum_{x_n} p(x_1, x_2, \ldots x_n)\right) =$$
$$= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} p(x_1, x_2, \ldots x_n) = 1$$

$$\sum_{x_j} p_j(x_j) = \sum_{x_j}\left(\sum_{x_1}\cdots\sum_{x_{j-1}}\sum_{x_{j+1}}\cdots\sum_{x_n} p(x_1, x_2, \ldots x_n)\right) =$$
$$= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} p(x_1, x_2, \ldots x_n) = 1$$

$$(5)$$
$$\sum_{x_i}\sum_{x_j} p_{ij}(x_i, x_j) = \sum_{x_i}\sum_{x_j}\left(\sum_{x_1}\cdots\sum_{x_{i-1}}\sum_{x_{i+1}}\cdots\sum_{x_{j-1}}\sum_{x_{j+1}}\cdots\sum_{x_n} p(x_1, x_2, \ldots x_n)\right) =$$
$$= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} p(x_1, x_2, \ldots x_n) = 1$$

Lets represent single-node marginal probabilities, using pairwise marginal probabilities (keeping in mind Equation (3)):

$$p_i(x_i) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} p(x_1, x_2, \ldots x_n) =$$

$$= \sum_{x_j} \left(\sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p(x_1, x_2, \ldots x_n) \right) =$$

$$= \sum_{x_j} p_{ij}(x_i, x_j)$$

(6)

$$p_j(x_j) = \sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p(x_1, x_2, \ldots x_n) =$$

$$= \sum_{x_i} \left(\sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p(x_1, x_2, \ldots x_n) \right) =$$

$$= \sum_{x_i} p_{ij}(x_i, x_j)$$

Lets consider summation over all edges - of some functions $f_i(x_i)$ and $f_j(x_j)$ (these functions 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:

(7)
$$\sum_{(i,j)} f_i(x_i) = \sum_i (q_i \times f_i(x_i))$$

$$\sum_{(i,j)} f_j(x_j) = \sum_j (q_j \times f_j(x_j))$$

Hammersley-Clifford Theorem [2, 3, 4] 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:

(8) $\quad p(x_1, x_2, \ldots x_n) = \dfrac{1}{Z} \times \left(\prod_{(i,j)} \psi_{ij}(x_i, x_j) \right) \times \left(\prod_i \phi_i(x_i) \right) \times \left(\prod_j \phi_j(x_j) \right)$

Boltzmann's Law:

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

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

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

From (8) and (9):

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

2

So:

$$(10) \quad E(x_1, x_2, \ldots x_n) = -\left(\sum_{(i,j)} ln(\psi_{ij}(x_i, x_j))\right) - \left(\sum_i ln(\phi_i(x_i))\right) - \left(\sum_j ln(\phi_j(x_j))\right)$$

Lets, assume that we have function $b(x_1, x_2, \ldots x_n)$, which approximates real probability density function $p(x_1, x_2, \ldots x_n)$. It means, that Equations (1)–(6) the same for $b(x_1, x_2, \ldots x_n)$:

$$(11) \quad \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1, x_2, \ldots x_n) = 1$$

$$(12) \quad
\begin{aligned}
b_i(x_i) &= \sum_{x_1}\cdots \sum_{x_{i-1}}\sum_{x_{i+1}}\cdots\sum_{x_n} b(x_1, x_2, \ldots x_n) \\
b_j(x_i) &= \sum_{x_1}\cdots \sum_{x_{j-1}}\sum_{x_{j+1}}\cdots\sum_{x_n} b(x_1, x_2, \ldots x_n)
\end{aligned}$$

$$(13) \quad b_{ij}(x_i, x_j) = \sum_{x_1}\cdots \sum_{x_{i-1}}\sum_{x_{i+1}}\cdots \sum_{x_{j-1}}\sum_{x_{j+1}}\cdots\sum_{x_n} b(x_1, x_2, \ldots x_n)$$

$$(14) \quad
\begin{aligned}
\sum_{x_i} b_i(x_i) &= 1 \\
\sum_{x_j} b_j(x_j) &= 1
\end{aligned}$$

$$(15) \quad \sum_{x_i}\sum_{x_j} b_{ij}(x_i, x_j) = 1$$

$$(16) \quad
\begin{aligned}
b_i(x_i) &= \sum_{x_j} b_{ij}(x_i, x_j) \\
b_j(x_j) &= \sum_{x_i} b_{ij}(x_i, x_j)
\end{aligned}$$

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

$$(17) \quad
\begin{aligned}
D_{KL}(b||p) &= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(b(x_1, x_2, \ldots x_n) \times ln\left(\frac{b(x_1, x_2, \ldots x_n)}{p(x_1, x_2, \ldots x_n)}\right)\right) = \\
&= \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1, x_2, \ldots x_n) \times ln\left(b(x_1, x_2, \ldots x_n)\right)\right) - \\
&\quad - \left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} b(x_1, x_2, \ldots x_n) \times ln\left(p(x_1, x_2, \ldots x_n)\right)\right)
\end{aligned}$$

Lest substitute $p(x_1, x_2, \ldots x_n)$ using expression from equation (9):

$$D_{KL}(b||p) = \left(\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \times ln\left(b(x_1, x_2, \ldots x_n)\right) \right) -$$

$$- \left(\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \times ln\left(\frac{1}{Z} \times \exp\left(-E(x_1, x_2, \ldots x_n)\right) \right) \right) =$$

$$= \left(\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \times ln\left(b(x_1, x_2, \ldots x_n)\right) \right) -$$

(18)
$$- \left(\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \times \left(-ln(Z) - E(x_1, x_2, \ldots x_n)\right) \right) =$$

$$= \left(\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \times ln\left(b(x_1, x_2, \ldots x_n)\right) \right) +$$

$$+ \left(\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \times ln(Z) \right) +$$

$$+ \left(\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \times E(x_1, x_2, \ldots x_n) \right)$$

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:

$$D_{KL}(b||p) = \left(\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \times ln\left(b(x_1, x_2, \ldots x_n)\right) \right) +$$

(19)
$$+ \left(\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \right) \times ln(Z) +$$

$$+ \left(\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \times E(x_1, x_2, \ldots x_n) \right)$$

Taking into account equation (11), we can substitute multiplier near $ln(Z)$ – to 1:

$$D_{KL}(b||p) = \left(\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \times ln\left(b(x_1, x_2, \ldots x_n)\right) \right) +$$

(20)
$$+ \left(\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \times E(x_1, x_2, \ldots x_n) \right) + ln(Z)$$

Lets use following notation:

$$U(b(x_1, x_2, \ldots x_n)) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} \left(b(x_1, x_2, \ldots x_n) \times E(x_1, x_2, \ldots x_n)\right)$$

(21)
$$-H(b(x_1, x_2, \ldots x_n)) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} \left(b(x_1, x_2, \ldots x_n) \times ln\left(b(x_1, x_2, \ldots x_n)\right)\right)$$

$$G(b(x_1, x_2, \ldots x_n)) = U(b(x_1, x_2, \ldots x_n)) - H(b(x_1, x_2, \ldots x_n))$$
$$-F(b(x_1, x_2, \ldots x_n)) = -F = ln(Z)$$

So:

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

Let's transform $U(b(x_1, x_2, \ldots x_n))$, using expression for energy (10):

(23)

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

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

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

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

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

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

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

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

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

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

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

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:

(24)

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

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

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

Taking into account Equations (12) and (13): we can substitute summations of probability density function – to marginal probabilities:

(25)
$$U(b(x_1, x_2, \ldots x_n)) = - \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} ln(\psi_{ij}(x_i, x_j)) \times b_{ij}(x_i, x_j) \right) -$$

$$- \left(\sum_i \sum_{x_i} ln(\phi_i(x_i)) \times b_i(x_i) \right) - \left(\sum_j \sum_{x_j} ln(\phi_j(x_j)) \times b_j(x_j) \right)$$

Lets introduce following variables ("local energies"):

(26)
$$E_i(x_i) = -ln(\phi_i(x_i))$$
$$E_j(x_j) = -ln(\phi_j(x_j))$$
$$E_{ij}(x_i, x_j) = -ln(\psi_{ij}(x_i, x_j)) - ln(\phi_i(x_i)) - ln(\phi_j(x_j))$$

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" (26):

(27)
$$ln(\phi_i(x_i)) = -E_i(x_i)$$
$$ln(\phi_j(x_j)) = -E_j(x_j)$$
$$ln(\psi_{ij}(x_i, x_j)) = -E_{ij}(x_i, x_j) + E_i(x_i) + E_j(x_j)$$

Lets transform $U(b(x_1, x_2, \ldots x_n))$ using Equations (27):

(28)
$$U(b(x_1, x_2, \ldots x_n)) = - \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times (-E_{ij}(x_i, x_j) + E_i(x_i) + E_j(x_j)) \right) -$$

$$- \left(\sum_i \sum_{x_i} b_i(x_i) \times (-E_i(x_i)) \right) - \left(\sum_j \sum_{x_j} b_j(x_j) \times (-E_j(x_j)) \right) =$$

$$= \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times E_{ij}(x_i, x_j) \right) - \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times E_i(x_i) \right) -$$

$$- \left(\sum_{(i,j)} \sum_{x_j} \sum_{x_i} b_{ij}(x_i, x_j) \times E_j(x_j) \right) + \left(\sum_i \sum_{x_i} b_i(x_i) \times E_i(x_i) \right) + \left(\sum_j \sum_{x_j} b_j(x_j) \times E_j(x_j) \right)$$

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:

(29)
$$U(b(x_1, x_2, \ldots x_n)) = \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times E_{ij}(x_i, x_j) \right) - \left(\sum_{(i,j)} \sum_{x_i} E_i(x_i) \times \sum_{x_j} b_{ij}(x_i, x_j) \right) -$$

$$- \left(\sum_{(i,j)} \sum_{x_j} E_j(x_j) \times \sum_{x_i} b_{ij}(x_i, x_j) \right) + \left(\sum_i \sum_{x_i} b_i(x_i) \times E_i(x_i) \right) + \left(\sum_j \sum_{x_j} b_j(x_j) \times E_j(x_j) \right)$$

Taking into account (16) – we can substitute summations over two-node marginal probabilities – to single-node marginal probabilities:

(30)
$$U(b(x_1, x_2, \ldots x_n)) = \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times E_{ij}(x_i, x_j) \right) - \left(\sum_{(i,j)} \sum_{x_i} E_i(x_i) \times b_i(x_i) \right) -$$

$$- \left(\sum_{(i,j)} \sum_{x_j} E_j(x_j) \times b_j(x_j) \right) + \left(\sum_i \sum_{x_i} b_i(x_i) \times E_i(x_i) \right) + \left(\sum_j \sum_{x_j} b_j(x_j) \times E_j(x_j) \right)$$

Finally, we can substitute summations over edges – to summation over nodes, as described by Equation (7):

(31)
$$U(b(x_1, x_2, \ldots x_n)) = \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times E_{ij}(x_i, x_j) \right) - \left(\sum_{x_i} \sum_{(i,j)} E_i(x_i) \times b_i(x_i) \right) -$$

$$- \left(\sum_{x_j} \sum_{(i,j)} E_j(x_j) \times b_j(x_j) \right) + \left(\sum_i \sum_{x_i} b_i(x_i) \times E_i(x_i) \right) + \left(\sum_j \sum_{x_j} b_j(x_j) \times E_j(x_j) \right) =$$

$$= \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times E_{ij}(x_i, x_j) \right) - \left(\sum_{x_i} \sum_i q_i \times E_i(x_i) \times b_i(x_i) \right) -$$

$$- \left(\sum_{x_j} \sum_j q_j \times E_j(x_j) \times b_j(x_j) \right) + \left(\sum_i \sum_{x_i} b_i(x_i) \times E_i(x_i) \right) + \left(\sum_j \sum_{x_j} b_j(x_j) \times E_j(x_j) \right) =$$

$$= \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times E_{ij}(x_i, x_j) \right) -$$

$$- \left(\sum_i \sum_{x_i} (q_i - 1) \times E_i(x_i) \times b_i(x_i) \right) - \left(\sum_j \sum_{x_j} (q_j - 1) \times E_j(x_j) \times b_j(x_j) \right)$$

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

energies":

$$U(b(x_1, x_2, \dots x_n)) = \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times E_{ij}(x_i, x_j) \right) -$$

(32)

$$- \left(\sum_i \sum_{x_i} (q_i - 1) \times E_i(x_i) \times b_i(x_i) \right) - \left(\sum_j \sum_{x_j} (q_j - 1) \times E_j(x_j) \times b_j(x_j) \right)$$

Now, lets transform $-H(b(x_1, x_2, \dots x_n))$ from Equations (21):

(33) $\quad -H(b(x_1, x_2, \dots x_n)) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} (b(x_1, x_2, \dots x_n) \times ln\,(b(x_1, x_2, \dots x_n)))$

Lets introduce additional statement [5] (which can be proven by induction [6]) – 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:

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

So, lets rewrite Equation (33), keeping in mind Equation (34):

(35)

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

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

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

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

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

$$- \left(\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \times \sum_i ln\left(b_i(x_i)^{q_i - 1} \right) \right) - \left(\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \times \sum_j ln\left(b_j(x_j)^{q_j - 1} \right) \right) =$$

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

$$- \left(\sum_i \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \times ln\left(b_i(x_i)^{q_i - 1} \right) \right) - \left(\sum_j \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} b(x_1, x_2, \dots x_n) \times ln\left(b_j(x_j)^{q_j - 1} \right) \right)$$

Lets rearrange sums, according to dependance of marginal probabilities:

(36)

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

$$- \left(\sum_i \sum_{x_i} \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \times ln\left(b_i(x_i)^{q_i-1}\right) \right) -$$

$$- \left(\sum_j \sum_{x_j} \sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \times ln\left(b_j(x_j)^{q_j-1}\right) \right) =$$

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

$$- \left(\sum_i \sum_{x_i} ln\left(b_i(x_i)^{q_i-1}\right) \times \left(\sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \right) \right) -$$

$$- \left(\sum_j \sum_{x_j} ln\left(b_j(x_j)^{q_j-1}\right) \times \left(\sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \right) \right)$$

Again, taking into account Equations (12) and (13): we can substitute summations of probability density function – to marginal probabilities:

(37)

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

$$- \left(\sum_i \sum_{x_i} ln\left(b_i(x_i)^{q_i-1}\right) \times \left(\sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \right) \right) -$$

$$- \left(\sum_j \sum_{x_j} ln\left(b_j(x_j)^{q_j-1}\right) \times \left(\sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} b(x_1, x_2, \ldots x_n) \right) \right) =$$

$$= \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} ln\left(b_{ij}(x_i, x_j)\right) \times b_{ij}(x_i, x_j) \right) -$$

$$- \left(\sum_i \sum_{x_i} ln\left(b_i(x_i)^{q_i-1}\right) \times b_i(x_i) \right) - \left(\sum_j \sum_{x_j} ln\left(b_j(x_j)^{q_j-1}\right) \times b_j(x_j) \right)$$

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

(38)
$$- H(b(x_1, x_2, \ldots x_n)) = \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} ln\left(b_{ij}(x_i, x_j) \right) \times b_{ij}(x_i, x_j) \right) -$$
$$- \left(\sum_i \sum_{x_i} (q_i - 1) \times ln(b_i(x_i)) \times b_i(x_i) \right) - \left(\sum_j \sum_{x_j} (q_j - 1) \times ln(b_j(x_j)) \times b_j(x_j) \right)$$

Lets represent Kullback-Leibler divergence (22) – in terms of marginal probabilities and local energies (keeping in mind Equations (32) and (38)):

(39)
$$D_{KL}(b||p) = U(b(x_1, x_2, \ldots x_n)) - H(b(x_1, x_2, \ldots x_n)) + ln(Z) =$$

$$= \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times E_{ij}(x_i, x_j) \right) -$$

$$- \left(\sum_i \sum_{x_i} (q_i - 1) \times E_i(x_i) \times b_i(x_i) \right) - \left(\sum_j \sum_{x_j} (q_j - 1) \times E_j(x_j) \times b_j(x_j) \right) +$$

$$+ \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} ln\left(b_{ij}(x_i, x_j) \right) \times b_{ij}(x_i, x_j) \right) -$$

$$- \left(\sum_i \sum_{x_i} (q_i - 1) \times ln(b_i(x_i)) \times b_i(x_i) \right) - \left(\sum_j \sum_{x_j} (q_j - 1) \times ln(b_j(x_j)) \times b_j(x_j) \right) + ln(Z) =$$

$$= \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times (E_{ij}(x_i, x_j) + ln\left(b_{ij}(x_i, x_j) \right)) \right) -$$

$$- \left(\sum_i \sum_{x_i} (q_i - 1) \times b_i(x_i) \times (E_i(x_i) + ln(b_i(x_i))) \right) - \left(\sum_j \sum_{x_j} (q_j - 1) \times b_j(x_j) \times (E_j(x_j) + ln(b_j(x_j))) \right) + ln(Z)$$

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:

(40)
$$D_{KL}(b||p) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} \left(b(x_1, x_2, \ldots x_n) \times ln\left(\frac{b(x_1, x_2, \ldots x_n)}{p(x_1, x_2, \ldots x_n)} \right) \right) =$$

$$= \left(\sum_{(i,j)} \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \times (E_{ij}(x_i, x_j) + ln\left(b_{ij}(x_i, x_j) \right)) \right) -$$

$$- \left(\sum_i \sum_{x_i} (q_i - 1) \times b_i(x_i) \times (E_i(x_i) + ln(b_i(x_i))) \right) - \left(\sum_j \sum_{x_j} (q_j - 1) \times b_j(x_j) \times (E_j(x_j) + ln(b_j(x_j))) \right) + ln(Z)$$

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 (14)–(16).

Lets use Lagrange multipliers method [8] for this purpose. Lets construct lagrangian, according to restrictions from Equations (14)–(16):

$$\mathcal{L} = D_{KL}(b||p) + \left(\sum_i \gamma_i \times \left(1 - \sum_{x_i} b_i(x_i) \right) \right) + \left(\sum_j \gamma_j \times \left(1 - \sum_{x_j} b_j(x_j) \right) \right) +$$

$$(41) \quad + \left(\sum_{(i,j)} \gamma_{ij} \times \left(1 - \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) \right) \right) + \left(\sum_i \sum_{x_i} \sum_{j \in N(i)} \lambda_{ji}(x_i) \times \left(b_i(x_i) - \sum_{x_j} b_{ij}(x_i, x_j) \right) \right) +$$

$$+ \left(\sum_j \sum_{x_j} \sum_{i \in N(j)} \lambda_{ij}(x_j) \times \left(b_j(x_j) - \sum_{x_i} b_{ij}(x_i, x_j) \right) \right)$$

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 $\mathcal{L}$:

$$(42) \quad \frac{\partial \mathcal{L}}{\partial b_{ij}(x_i, x_j)} = 0, \ (\forall (i,j) \in Edges, \ \forall x_i \in StatesOfNode(i), \ \forall x_j \in StatesOfNode(j))$$

$$(43) \quad \frac{\partial \mathcal{L}}{\partial b_i(x_i)} = 0, \ (\forall i \in OddNodes, \ \forall x_i \in StatesOfNode(i))$$

$$(44) \quad \frac{\partial \mathcal{L}}{\partial b_j(x_j)} = 0, \ (\forall j \in EvenNodes, \ \forall x_j \in StatesOfNode(j))$$

Lets transform Equation (42), using Equation (41):

$$\frac{\partial \mathcal{L}}{\partial b_{ij}(x_i, x_j)} = \frac{\partial \left(D_{KL}(b||p) \right)}{\partial b_{ij}(x_i, x_j)} + \frac{\partial \left(- \gamma_{ij} \times b_{ij}(x_i, x_j) \right)}{\partial b_{ij}(x_i, x_j)} +$$

$$(45) \quad + \frac{\partial \left(- \lambda_{ji}(x_i) \times b_{ij}(x_i, x_j) \right)}{\partial b_{ij}(x_i, x_j)} + \frac{\partial \left(- \lambda_{ij}(x_j) \times b_{ij}(x_i, x_j) \right)}{\partial b_{ij}(x_i, x_j)} =$$

$$= \frac{\partial \left(D_{KL}(b||p) \right)}{\partial b_{ij}(x_i, x_j)} - \gamma_{ij} - \lambda_{ji}(x_i) - \lambda_{ij}(x_j) = 0$$

11

For the further transformation of partial derivative – lets use expression for $D_{KL}(b||p)$ from Equation (40):

$$\frac{\partial \mathcal{L}}{\partial b_{ij}(x_i, x_j)} = \frac{\partial \left(D_{KL}(b||p) \right)}{\partial b_{ij}(x_i, x_j)} - \gamma_{ij} - \lambda_{ji}(x_i) - \lambda_{ij}(x_j) =$$

(46)

$$= \frac{\partial \left(b_{ij}(x_i, x_j) \times \left(E_{ij}(x_i, x_j) + ln\left(b_{ij}(x_i, x_j) \right) \right) \right)}{\partial b_{ij}(x_i, x_j)} - \gamma_{ij} - \lambda_{ji}(x_i) - \lambda_{ij}(x_j) =$$

$$= E_{ij}(x_i, x_j) + \frac{\partial \left(b_{ij}(x_i, x_j) \times ln\left(b_{ij}(x_i, x_j) \right) \right)}{\partial b_{ij}(x_i, x_j)} - \gamma_{ij} - \lambda_{ji}(x_i) - \lambda_{ij}(x_j) =$$

$$= E_{ij}(x_i, x_j) + ln\left(b_{ij}(x_i, x_j) \right) + 1 - \gamma_{ij} - \lambda_{ji}(x_i) - \lambda_{ij}(x_j) = 0$$

So, finally, from Equation (46) – we can derive expression for $ln\left(b_{ij}(x_i, x_j) \right)$:

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

Now, lets transform Equation (43), using Equation (41):

(48)

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

$$= \frac{\partial \left(D_{KL}(b||p) \right)}{\partial b_i(x_i)} - \gamma_i + \sum_{j \in N(i)} \left(\frac{\partial \left(\lambda_{ji}(x_i) \times b_i(x_i) \right)}{\partial b_i(x_i)} \right) =$$

$$= \frac{\partial \left(D_{KL}(b||p) \right)}{\partial b_i(x_i)} - \gamma_i + \sum_{j \in N(i)} \lambda_{ji}(x_i) = 0$$

Again, for the further transformation of partial derivative – lets use expression for $D_{KL}(b||p)$ from Equation (40):

(49)

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

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

$$= -(q_i - 1) \times E_i(x_i) - (q_i - 1) \times \left(1 + ln(b_i(x_i)) \right) - \gamma_i + \sum_{j \in N(i)} \lambda_{ji}(x_i) = 0$$

So, finally, from Equation (49) – we can derive expression for $ln\,(b_i(x_i))$:

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

$$(50) \quad (1 + ln(b_i(x_i))) = -E_i(x_i) - \frac{\gamma_i}{q_i - 1} + \frac{\sum_{j \in N(i)} \lambda_{ji}(x_i)}{q_i - 1}$$

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

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

$$(51) \quad ln(b_j(x_j)) = -E_j(x_j) - \frac{\gamma_j}{q_j - 1} - 1 + \frac{1}{q_j - 1} \times \sum_{i \in N(j)} \lambda_{ij}(x_j)$$

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

$$(52) \quad \begin{cases} ln\,(b_{ij}(x_i, x_j)) = \gamma_{ij} + \lambda_{ji}(x_i) + \lambda_{ij}(x_j) - E_{ij}(x_i, x_j) - 1 \\[2ex] ln(b_i(x_i)) = -E_i(x_i) - \dfrac{\gamma_i}{q_i - 1} - 1 + \dfrac{1}{q_i - 1} \times \displaystyle\sum_{j \in N(i)} \lambda_{ji}(x_i) \\[3ex] ln(b_j(x_j)) = -E_j(x_j) - \dfrac{\gamma_j}{q_j - 1} - 1 + \dfrac{1}{q_j - 1} \times \displaystyle\sum_{i \in N(j)} \lambda_{ij}(x_j) \end{cases}$$

As you remember (from Equation (41)): 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:

$$(53) \quad \lambda_{ji}(x_i) = \sum_{k \in N(i) \backslash j} ln\,(m_{k \to i}(x_i))$$

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

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

$$(54) \quad \lambda_{ij}(x_j) = \sum_{k \in N(j) \backslash i} ln\,(m_{k \to j}(x_j))$$

13

Lets rewrite Equations (52), using Equations (53) and (54):

$$(55) \quad \begin{cases} ln\left(b_{ij}(x_i, x_j)\right) = -E_{ij}(x_i, x_j) + \gamma_{ij} - 1 + \left(\sum_{k \in N(i) \backslash j} ln\left(m_{k \rightarrow i}(x_i)\right)\right) + \left(\sum_{k \in N(j) \backslash i} ln\left(m_{k \rightarrow j}(x_j)\right)\right) \\[3em] ln(b_i(x_i)) = -E_i(x_i) - \dfrac{\gamma_i}{q_i - 1} - 1 + \dfrac{1}{q_i - 1} \times \sum_{j \in N(i)} \left(\sum_{k \in N(i) \backslash j} ln\left(m_{k \rightarrow i}(x_i)\right)\right) \\[3em] ln(b_j(x_j)) = -E_j(x_j) - \dfrac{\gamma_j}{q_j - 1} - 1 + \dfrac{1}{q_j - 1} \times \sum_{i \in N(j)} \left(\sum_{k \in N(j) \backslash i} ln\left(m_{k \rightarrow j}(x_j)\right)\right) \end{cases}$$

Also, it can be easy shown that:

$$(56) \quad \begin{aligned} \frac{1}{q_i - 1} \times \sum_{j \in N(i)} \sum_{k \in N(i) \backslash j} f(k) &\equiv \sum_{k \in N(i)} f(k) \\ \frac{1}{q_j - 1} \times \sum_{i \in N(j)} \sum_{k \in N(j) \backslash i} f(k) &\equiv \sum_{k \in N(j)} f(k) \end{aligned}$$

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

$$(57) \quad \begin{cases} ln\left(b_{ij}(x_i, x_j)\right) = -E_{ij}(x_i, x_j) + \gamma_{ij} - 1 + \left(\sum_{k \in N(i) \backslash j} ln\left(m_{k \rightarrow i}(x_i)\right)\right) + \left(\sum_{k \in N(j) \backslash i} ln\left(m_{k \rightarrow j}(x_j)\right)\right) \\[3em] ln(b_i(x_i)) = -E_i(x_i) - \dfrac{\gamma_i}{q_i - 1} - 1 + \sum_{k \in N(i)} ln\left(m_{k \rightarrow i}(x_i)\right) \\[3em] ln(b_j(x_j)) = -E_j(x_j) - \dfrac{\gamma_j}{q_j - 1} - 1 + \sum_{k \in N(j)} ln\left(m_{k \rightarrow j}(x_j)\right) \end{cases}$$

Lets substitute local energies $(E_i(x_i), E_j(x_j)$ and $E_{ij}(x_i, x_j))$ – using Equations (26):

$$(58) \quad \begin{cases} ln\left(b_{ij}(x_i, x_j)\right) = ln(\psi_{ij}(x_i, x_j)) + ln(\phi_i(x_i)) + ln(\phi_j(x_j)) + \gamma_{ij} - 1 + \\[1.5em] \quad + \left(\sum_{k \in N(i) \backslash j} ln\left(m_{k \rightarrow i}(x_i)\right)\right) + \left(\sum_{k \in N(j) \backslash i} ln\left(m_{k \rightarrow j}(x_j)\right)\right) \\[3em] ln(b_i(x_i)) = ln(\phi_i(x_i)) - \dfrac{\gamma_i}{q_i - 1} - 1 + \sum_{k \in N(i)} ln\left(m_{k \rightarrow i}(x_i)\right) \\[3em] ln(b_j(x_j)) = ln(\phi_j(x_j)) - \dfrac{\gamma_j}{q_j - 1} - 1 + \sum_{k \in N(j)} ln\left(m_{k \rightarrow j}(x_j)\right) \end{cases}$$

Now, lets exponentiate Equations (58) – to get expressions for marginal probabilities:

(59)
$$
\begin{cases}
b_{ij}(x_i, x_j) = \exp\left(\gamma_{ij} - 1\right) \times \left(\prod_{k \in N(i) \setminus j} m_{k \to i}(x_i) \right) \times \phi_i(x_i) \times \psi_{ij}(x_i, x_j) \times \phi_j(x_j) \times \left(\prod_{k \in N(j) \setminus i} m_{k \to j}(x_j) \right) \\[3ex]
b_i(x_i) = \exp\left(-\dfrac{\gamma_i}{q_i - 1} - 1 \right) \times \phi_i(x_i) \times \left(\prod_{k \in N(i)} m_{k \to i}(x_i) \right) \\[3ex]
b_j(x_j) = \exp\left(-\dfrac{\gamma_j}{q_j - 1} - 1 \right) \times \phi_j(x_j) \times \left(\prod_{k \in N(j)} m_{k \to j}(x_j) \right)
\end{cases}
$$

If we look at Equations (59) – we can notice, that following multipliers does not depend on states of graph nodes (x_i and x_j):

(60)
$$
\begin{cases}
\exp\left(\gamma_{ij} - 1\right) \neq f(x_i, x_j) \\[2ex]
\exp\left(-\dfrac{\gamma_i}{q_i - 1} - 1 \right) \neq f(x_i) \\[2ex]
\exp\left(-\dfrac{\gamma_j}{q_j - 1} - 1 \right) \neq f(x_j)
\end{cases}
$$

Lets denote:

(61)
$$
\begin{cases}
Z_{ij} = \exp\left(\gamma_{ij} - 1\right) \\[2ex]
Z_i = \exp\left(-\dfrac{\gamma_i}{q_i - 1} - 1 \right) \\[2ex]
Z_j = \exp\left(-\dfrac{\gamma_j}{q_j - 1} - 1 \right)
\end{cases}
$$

So, actually, it means, that given multipliers just corresponds to normalization coefficients (with respect to Equations (14) and (15)):

(62)
$$
\begin{cases}
b_{ij}(x_i, x_j) = Z_{ij} \times \left(\prod_{k \in N(i) \setminus j} m_{k \to i}(x_i) \right) \times \phi_i(x_i) \times \psi_{ij}(x_i, x_j) \times \phi_j(x_j) \times \left(\prod_{k \in N(j) \setminus i} m_{k \to j}(x_j) \right), \\
\qquad \text{where } Z_{ij} \text{ is such that } \sum_{x_i} \sum_{x_j} b_{ij}(x_i, x_j) = 1 \\[3ex]
b_i(x_i) = Z_i \times \phi_i(x_i) \times \left(\prod_{k \in N(i)} m_{k \to i}(x_i) \right), \\
\qquad \text{where } Z_i \text{ is such that } \sum_{x_i} b_i(x_i) = 1 \\[3ex]
b_j(x_j) = Z_j \times \phi_j(x_j) \times \left(\prod_{k \in N(j)} m_{k \to j}(x_j) \right), \\
\qquad \text{where } Z_j \text{ is such that } \sum_{x_j} b_j(x_j) = 1
\end{cases}
$$

References

[1] 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`

[2] 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`.

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

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

[5] 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`

[6] 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`

[7] 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`

[8] `http://en.wikipedia.org/wiki/Lagrange_multiplier`

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