Statistical mechanics, graphical models and message passing algorithms

Marc Lelarge*

! Draft !

August 26, 2016

Contents

*email: marc.lelarge@ens.fr

These lecture notes have been written for a course given at the summer school in advanced probability organized at Novosibirsk State University, Sobolev Institute of Mathematics in August 2016. They are self-contained and cover part of the material of [5, 7, 8]. These lecture notes shows how ideas from statistical mechanics, graphical models and message passing algorithms can be combined on a particular example: the monomer-dimer problem. During the lectures, I also covered the sum-product algorithm and the Bethe free energy in a general context. These sections will be added in a future version of these lecture notes...

1 Some definitions

We consider a connected multigraph $G = (V, E)$. We denote by $v(G)$ the cardinality of V: $v(G) = |V|$. We denote by the same symbol ∂v the set of neighbors of node $v \in V$ and the set of edges incident to v. Also, $\partial u \backslash v$ is the set of neighbors of u in G from which we removed v. A matching is encoded by a binary vector, called its incidence vector, $\mathbf{b} = (b_e,\ e \in E) \in \{0,1\}^E$ defined by $b_e = 1$ if and only if the edge e belongs to the matching. We have for all $v \in V$, $\sum_{e\in\partial v} b_e \leq 1$. The size of the matching is given by $\sum_e b_e$. We will also use the following notation $e \in \mathbf{b}$ to mean that $b_e = 1$, i.e. that the edge e is in the matching. For a finite graph G, we define the matching number of G as $\nu(G) = \max\{\sum_e b_e\}$ where the maximum is taken over matchings of G.

The matching polytope $M(G)$ of a graph G is defined as the convex hull of incidence vectors of matchings in G. We define the fractional matching polytope as

$$FM(G) = \left\{ \mathbf{x} \in \mathbb{R}^E |\ x_e \geq 0,\ \sum_{e\in\partial v} x_e \leq 1 \right\}. \tag{1}$$

We also define the fractional matching number $\nu^*(G) = \max_{\mathbf{x}\in FM(G)} \sum_e x_e \geq \nu(G)$. It is well-known that: $M(G) = FM(G)$ if and only if G is bipartite and in this case, we have $\nu(G) = \nu^*(G)$, see [11].

For a given graph G, we denote by $m_k(G)$ the number of matchings of size k in G ($m_0(G) = 1$). For a parameter $z > 0$, we define the matching generating function:

$$P_G(z) = \sum_{k=0}^{\nu(G)} m_k(G) z^k.$$

In statistical physics, the function $P_G(z)$ is called the partition function. We introduce the family of probability distributions on the set of matchings parametrised by a parameter $z > 0$ and called the Gibbs measures:

$$\mu_G^z(\mathbf{B} = \mathbf{b}) = \frac{z^{\sum_e b_e}}{P_G(z)}. \tag{2}$$

Conventions: to ligthen the notation, we will usually write $\mu_G^z(\mathbf{b})$ instead of $\mu_G^z(\mathbf{B} = \mathbf{b})$. Also for a subset S of the set of matchings, we use the standard notation in probability:

$$\mu_G^z(S) = \mu_G^z(\mathbf{B} \in S) = \sum_{\mathbf{b}\in S} \mu_G^z(\mathbf{b}).$$

For example, if $S = \{\mathbf{b},\ b_e = 1\}$ for a given $e \in E$, we write:

$$\mu_G^z(B_e = 1) = \sum_{\mathbf{b}\in S} \mu_G^z(\mathbf{b}).$$

Expectation with respect to μ_G^z is denoted by $\langle . \rangle$ so that for a function $f(\mathbf{b})$ from the set of matchings to $\mathbb{R}$, we have

$$\langle f(\mathbf{B})\rangle = \sum_{\mathbf{b}} f(\mathbf{b})\mu_G^z(\mathbf{b}).$$

Also our original graph G is unoriented, we introduce the set $\overrightarrow{E}$ of directed edges of G comprising two directed edges $u \to v$ and $v \to u$ for each undirected edge $uv \in E$. For $\overrightarrow{e} \in \overrightarrow{E}$, we denote by $-\overrightarrow{e}$ the edge with opposite direction. With a slight abuse of notation, we also denote by ∂v the set of incident edges to $v \in V$ directed towards v.

2 A (very) little bit of statistical physics and Markov random fields

We define the internal energy $U_G(z)$ and the canonical entropy $S_G(z)$ as:

$$\begin{aligned} U_G(z) &= -\sum_{e\in E} \mu_G^z(B_e = 1), \\ S_G(z) &= -\sum_{\mathbf{b}} \mu_G^z(\mathbf{b}) \ln \mu_G^z(\mathbf{b}). \end{aligned}$$

Note that we have:

$$\begin{aligned} U_G(z) &= -\langle \sum_e B_e \rangle, \\ S_G(z) &= \langle \ln \mu_G^z(\mathbf{B})\rangle. \end{aligned}$$

The free entropy $\Phi_G(z)$ is then defined by

$$\Phi_G(z) = -U_G(z)\ln z + S_G(z).$$

A more conventional notation in the statistical physics literature corresponds to an inverse temperature $\beta = \ln z$.

Lemma 1. *We have $\Phi_G(z) = \ln P_G(z)$ and $\Phi'_G(z) = -\frac{U_G(z)}{z}$.*

Proof left as an exercise.

Lemma 2. *The function $U_G(z)$ is strictly decreasing and mapping $[0,\infty)$ to $(-\nu(G), 0]$.*

Proof. We have $-U_G(z) = \sum_k k m_k(G) z^k / P_G(z)$ so that taking the derivative and multiplying by z, we get:

$$\begin{aligned} -z(U_G)'(z) &= \frac{\sum_k k^2 m_k(G) z^k}{P_G(z)} - \left(\frac{\sum_k k m_k(G) z^k}{P_G(z)}\right)^2 \\ &= \sum_k \left(k - \frac{\sum_\ell \ell m_\ell(G) z^\ell}{P_G(z)}\right)^2 \frac{m_k(G) z^k}{P_G(z)} > 0. \end{aligned}$$

□

Lemma 3. *For all $z > 1$, we have*

$$-U_G(z) \le \nu(G) \le -U_G(z) + |E|\frac{\ln 2}{\ln z}.$$

In particular, we have $\lim_{z\to\infty} -U_G(z) = \nu(G)$.

Proof. For any $z > 1$, we have thanks to previous lemma

$$\int_1^z \frac{-U_G(s)}{s} ds \le -U_G(z) \ln z.$$

Hence by Lemma 1, we have

$$\ln \frac{P_G(z)}{P_G(1)} \le -U_G(z) \ln z.$$

We have $P_G(1) = \sum_k m_k(G) \le 2^{|E|}$ and $P_G(z) \ge z^{\nu(G)}$, so that we get

$$\nu(G) \le -U_G(z) + |E|\frac{\ln 2}{\ln z}.$$

□

For a finite graph $G = (V, E)$, we define: for $v \in V$ and $h \in \mathbb{N}$, let $G_h(v) = (V_h(v), E_h(v))$ be the subgraph of G induced by vertices at distance at most h from v; $\partial V_h(v)$ is the set of edges of G connecting $V_h(v)$ and $V \backslash V_h(v)$. For a vector $\mathbf{b} = (b_e,\ e \in E)$, and a set $S \subset E$, we define $\mathbf{b}_S = (b_e,\ e \in S)$.

Lemma 4. *(Markov property) For any $h \ge 1$ and $v \in V$, we have for any $\mathbf{b} \in \{0,1\}^E$ with $\sum_{e\in\partial u} b_e \le 1$ for all $u \in V$,*

$$\begin{aligned}\mu_G^z\left(\mathbf{B}_{E_h(v)} = \mathbf{b}_{E_h(v)} | \mathbf{B}_{E\backslash E_h(v)} = \mathbf{b}_{E\backslash E_h(v)}\right) &= \mu_G^z\left(\mathbf{B}_{E_h(v)} = \mathbf{b}_{E_h(v)} | \mathbf{B}_{\partial V_h(v)} = \mathbf{b}_{\partial V_h(v)}\right)\\ &= \mu^z_{G_h^{\mathbf{b}}(v)}\left(\mathbf{B}_{E_h(v)} = \mathbf{b}_{E_h(v)}\right),\end{aligned}$$

with $G_h^{\mathbf{b}}(v)$ obtained from $G_h(v)$ by removing all vertices $u \in V_h(v)$ such that $e = (uw) \in \partial V_h(v)$ and $b_e = 1$.

Proof left as an exercise.

3 Local recursions on finite trees and finite graphs

We first consider the case where G is a tree, i.e. a connected graph without cycle. To make this assumption clear, we denote it by $T = (V, E)$ instead of G. For any directed edge $u \to v$, we define $T_{u\to v}$ as the subtree containing u and v and obtained from T by removing all incident edges to v except the edge uv. A simple computation shows that

$$\begin{aligned}\frac{\mu^z_{T_{u\to v}}(B_{uv=1})}{\mu^z_{T_{u\to v}}(B_{uv=0})} &= \frac{z\prod_{w'\in\partial u\backslash v}\mu^z_{T_{w'\to u}}(B_{wu}=0)}{\prod_{w'\in\partial u\backslash v}\mu^z_{T_{w'\to u}}(B_{wu}=0)+\sum_{w\in\partial u\backslash v}\mu^z_{T_{w\to u}}(B_{wu}=1)\prod_{w'\in\partial u\backslash\{v,w\}}\mu^z_{T_{w'\to u}}(B_{wu}=0)}\\ &= \frac{z}{1+\sum_{w\in\partial u\backslash v}\frac{\mu^z_{T_{w\to u}}(B_{wu}=1)}{\mu^z_{T_{w\to u}}(B_{wu}=0)}}.\end{aligned}$$

We define:

$$Y_{u\to v}(z) = \frac{\mu_{T^z_{u\to v}(B_{uv=1})}}{\mu_{T^z_{u\to v}(B_{uv=0})}},$$

so that for a finite tree, we have:

$$Y_{u\to v}(z) = \frac{z}{1+\sum_{w\in\partial u\backslash v}Y_{w\to u}(z)}. \tag{3}$$

Then a simple computation shows that

$$\frac{\mu^z_T(B_{uv=1})}{\mu^z_T(B_{uv=0})} = \frac{Y_{u\to v}(z)Y_{v\to u}(z)}{z}. \tag{4}$$

Also, we gave an interpretation for the local recursion (3) for trees, only, we can define it for any graph G. Given a set of 'messages' $\mathbf{a}\in[0,\infty)^{\overrightarrow{E}}$, we define a new set of 'messages' $\mathbf{b}\in[0,\infty)^{\overrightarrow{E}}$ by:

$$b_{u\to v} = \frac{1}{1+\sum_{w\in\partial u\backslash v}a_{w\to u}}, \tag{5}$$

with the convention that the sum over the empty set equals zero. We denote by $\mathcal{R}_G$ the mapping sending $\mathbf{a}\in[0,\infty)^{\overrightarrow{E}}$ to $\mathbf{b}=\mathcal{R}_G(\mathbf{a})$. For $\overrightarrow{e}\in\overrightarrow{E}$, we also denote by $\mathcal{R}_{\overrightarrow{e}}:[0,\infty)^{\overrightarrow{E}}\to[0,\infty)$ the local update rule (5): $b_{\overrightarrow{e}}=\mathcal{R}_{\overrightarrow{e}}(\mathbf{a})$. Note that $\mathcal{R}_{u\to v}$ indeed depends only on messages on edges in $\partial u\backslash v$. We also denote by $z\mathcal{R}_G$ the mapping multiplying by z each component of the output of the mapping $\mathcal{R}_G$ (making the notation consistent).

Proposition 5. *(i) For any finite graph G and $z>0$, the fixed point equation:*

$$\mathbf{y} = z\mathcal{R}_G(\mathbf{y}) \tag{6}$$

has a unique attractive solution denoted $\mathbf{y}(z)\in(0,z)^{\overrightarrow{E}}$.

(ii) If in addition, G is a finite tree, then for all $e\in E$, the law of B_e under μ^z_G is a Bernoulli distribution with

$$\mu^z_G(B_e=1) = \frac{y_{\overrightarrow{e}}(z)y_{-\overrightarrow{e}}(z)}{z+y_{\overrightarrow{e}}(z)y_{-\overrightarrow{e}}(z)}. \tag{7}$$

Comparisons between vectors are always componentwise. Note that the last point follows directly from (4). Before proving this proposition, let define for all $v \in V$, the following function of the messages $(y_{\overrightarrow{e}},\ \overrightarrow{e} \in \partial v)$,

$$\mathcal{D}_v(\mathbf{y}) = \sum_{\overrightarrow{e} \in \partial v} \frac{y_{\overrightarrow{e}} \mathcal{R}_{-\overrightarrow{e}}(\mathbf{y})}{1 + y_{\overrightarrow{e}} \mathcal{R}_{-\overrightarrow{e}}(\mathbf{y})} \tag{8}$$

$$= \frac{\sum_{\overrightarrow{e} \in \partial v} y_{\overrightarrow{e}}}{1 + \sum_{\overrightarrow{e} \in \partial v} y_{\overrightarrow{e}}}. \tag{9}$$

Note that if the graph G is a tree, $\mathcal{D}_v(\mathbf{y}(z))$ is simply the probability for vertex v to be covered by a matching distributed according to μ_G^z.

Proof. For the first point, we follow the proof of Theorem 3 in [10]. Let $z > 0$ and define the sequence of messages: $\mathbf{x}^0(z) = 0$ and for $t \geq 0$,

$$x_{u \to v}^{t+1}(z) = \frac{z}{1 + \sum_{w \in \partial u \setminus v} x_{w \to u}^t(z)}. \tag{10}$$

The sequence $\mathbf{x}^{2t}(z)$ (resp. $\mathbf{x}^{2t+1}(z)$) is non-decreasing (resp. non-increasing). We define $\lim_{t\to\infty} \uparrow \mathbf{x}^{2t}(z) = \mathbf{x}^-(z)$ and $\lim_{t\to\infty} \downarrow \mathbf{x}^{2t+1}(z) = \mathbf{x}^+(z)$. For any $\mathbf{y}(z)$ fixed point of (6), a simple induction shows that

$$0 \leq \mathbf{x}^{2t}(z) \leq \mathbf{x}^-(z) \leq \mathbf{y}(z) \leq \mathbf{x}^+(z) \leq \mathbf{x}^{2t+1}(z) \leq z.$$

We now prove that $\mathbf{x}^-(z) = \mathbf{x}^+(z)$ finishing the proof of the first point. Note that we have $\mathbf{x}^+(z) = z\mathcal{R}_G(\mathbf{x}^-(z))$ and $\mathbf{x}^-(z) = z\mathcal{R}_G(\mathbf{x}^+(z))$. In particular for any $z > 0$, we have $x^+_{\overrightarrow{e}}(z)\mathcal{R}_{-\overrightarrow{e}}(\mathbf{x}^+(z)) = x^-_{-\overrightarrow{e}}(z)\mathcal{R}_{\overrightarrow{e}}(\mathbf{x}^-(z))$ so that in view of (8), we have

$$\sum_{v \in V} \mathcal{D}_v(\mathbf{x}^+(z)) = \sum_{v \in V} \mathcal{D}_v(\mathbf{x}^-(z)). \tag{11}$$

We see from (9) that for each $v \in V$, $\mathcal{D}_v(\mathbf{x})$ is an increasing function of the $\sum_{\overrightarrow{e} \in \partial v} x_{\overrightarrow{e}}$, so that (11) together with $\mathbf{x}^-(z) \leq \mathbf{x}^+(z)$ imply the desired result. $\square$

For $z > 0$, let $\mathbf{x}(z) \in \mathbb{R}^E$ be defined by

$$x_e(z) = \frac{y_{\overrightarrow{e}}(z) y_{-\overrightarrow{e}}(z)}{z + y_{\overrightarrow{e}}(z) y_{-\overrightarrow{e}}(z)} \in (0, 1), \tag{12}$$

where $\mathbf{y} = (y_{\overrightarrow{e}}, \overrightarrow{e} \in \overrightarrow{E})$ is the solution defined above.

We now give a reparametrization of the Gibbs distribution. For any vector $\mathbf{b} \in \{0,1\}^E$, we denote by $\mathbf{b}_{\partial v} \in \{0,1\}^{\partial v}$ its restriction to components in ∂v. We first define the marginal probabilities for $\mathbf{b}_{\partial v}$ such that $\sum_{e \in \partial v} b_e \leq 1$,

$$\mu_{\partial v}(\mathbf{b}_{\partial v}) = \left(1 - \sum_{e \in \partial v} x_e(z)\right)^{1 - \sum_{e \in \partial v} b_e} \prod_{e \in \partial v} x_e(z)^{b_e}, \tag{13}$$

and for $be \in \{0, 1\}$,

$$\mu_e(b_e) = x_e(z)^{b_e}(1 - x_e(z))^{1-b_e}, \tag{14}$$

where $x_e(z)$ is defined by (12).

Exercise 6. *Show that if G is a tree, these correspond to the marginals of μ_G^z.*

Given a graph $G = (V, E)$ and some set $F \subset E$, we define $d_F(v)$ as the degree of node v in the subgraph induced by F. A generalized loop is any subset F such that $d_F(v) \neq 1$ for all $v \in V$. We define $V(F)$ as the number of vertices covered by F, i.e. vertices with $d_F(v) \geq 1$.

Theorem 7. *For any graph G, we have for $z > 0$,*

$$\mu_G^z(\mathbf{b}) = \frac{1}{1+L} \frac{\prod_{v \in V} \mu_{\partial v}(\mathbf{b}_{\partial v})}{\prod_{e \in E} \mu_e(b_e)}, \tag{15}$$

with

$$L = \sum_{\emptyset \neq F \subset E} (-1)^{V(F)} \prod_{v \in V} (d_F(v) - 1) \prod_{e \in F} \frac{x_e(z)}{1 - x_e(z)}, \tag{16}$$

where only generalized loops F lead to a non-zero term L.

Note in particular that if G is a tree, then $L = 0$.

Proof. The fact that μ_G^z can be written as (15) (called tree-based reparameterization in [13]) follows from a direct application of the definitions.

Lemma 8. *For any $v \in V$, $z > 0$, we have*

$$\frac{\mu_{\partial v}(\mathbf{b}_{\partial v})}{\prod_{e \in \partial v} \mu_e(b_e)} = 1 - \sum_{S \subset \partial v} (-1)^{|S|} (|S| - 1) \prod_{e \in S} \frac{b_e - x_e(z)}{1 - x_e(z)}.$$

To simplify notation, we write in the proofs x_e instead of $x_e(z)$.

Proof. Note that if $b_f = 1$, the left-hand side is equal to $\prod_{e \neq f}(1 - x_e)^{-1}$, while if $\sum_{e \in \partial v} b_e = 0$, it is equal to $\frac{1 - \sum_{e \in \partial v} x_e}{\prod_{e \in \partial v}(1 - x_e)}$. We need to check that the right-hand side agrees in these two cases (which are the only possible ones due to the constraint of being a matching).

Let consider the case $b_f = 1$, then the right-hand side (denoted R) equals:

$$\begin{aligned}
R &= 1 - \sum_{|S| \geq 1, f \notin S} (-1)^{|S|} (|S| - 1) \prod_{e \in S} \frac{-x_e}{1 - x_e} \\
&\quad - \sum_{|S| \geq 1, f \in S} (-1)^{|S|} (|S| - 1) \prod_{e \in S, e \neq f} \frac{-x_e}{1 - x_e} \\
&= 1 - \sum_{|S| \geq 1, f \notin S} (-1)^{|S|+1} \prod_{e \in S} \frac{-x_e}{1 - x_e} \\
&= 1 + \sum_{|S| \geq 1, f \notin S} \frac{\prod_{e \in S} x_e \prod_{e' \notin S, e' \neq f}(1 - x_{e'})}{\prod_{e \neq f}(1 - x_e)} \\
&= \frac{1}{\prod_{e \neq f}(1 - x_e)}.
\end{aligned}$$

A similar computation shows the second case. □

We now compute L. By definition, we have

$$1+L=\sum_{\mathbf{b}}\prod_{e}\mu_e(b_e)\prod_{v}\frac{\mu_{\partial v}(\mathbf{b}_{\partial v})}{\prod_{e\in\partial v}\mu_e(b_e)}.$$

By Lemma 8, we have

$$\begin{aligned}P &:= \prod_{v}\frac{\mu_{\partial v}(\mathbf{b}_{\partial v})}{\prod_{e\in\partial v}\mu_e(b_e)}\\ &= \prod_{v}\left(1+\sum_{S\subset\partial v}(-1)^{|S|-1}\left(|S|-1\right)\prod_{e\in S}\frac{b_e-x_e}{1-x_e}\right)\end{aligned}$$

$1+L$ can be seen as an expectation of P where the B_e are independent Bernoulli random variables with parameter x_e. In particular expanding P, we see that only the terms $(B_e-x_e)^2$ will contribute to its expectation so that we get

$$\begin{aligned}L &= \sum_{\emptyset\neq F\subset E}\prod_{v}\left(\frac{(-1)^{d_F(v)-1}(d_F(v)-1)}{\prod_{e\in\partial v\cap F}(1-x_e)}\right)\prod_{e\in F}x_e(1-x_e)\\ &= \sum_{\emptyset\neq F\subset E}(-1)^{V(F)}\prod_{v}(d_F(v)-1)\prod_{e\in F}\frac{x_e}{1-x_e},\end{aligned}$$

where in the last claim, we used $\prod_v(-1)^{d_F(v)}=1$.

□

Hence, summarizing our results so far, we have if G is a tree:

$$\begin{aligned}U_G(z) &= -\sum_{e\in E}x_e(z),\\ S_G(z) &= -\sum_{\mathbf{b}}\mu_G^z(\mathbf{b})\ln\mu_G^z(\mathbf{b})\\ &= -\sum_{\mathbf{b}}\mu_G^z(\mathbf{b})\left(\sum_{v}\ln\mu_{\partial v}(\mathbf{b}_{\partial v})-\sum_{e}\ln\mu_e(b_e)\right)\\ &= -\sum_{v}\sum_{\mathbf{b}_{\partial v}\in\{0,1\}^{\partial v}}\mu_{\partial v}(\mathbf{b}_{\partial v})\ln\mu_{\partial v}(\mathbf{b}_{\partial v})+\sum_{e}\sum_{b_e\in\{0,1\}}\mu_e(b_e)\ln\mu_e(b_e)\\ &= \frac{1}{2}\sum_{v\in V}\left\{\sum_{e\in\partial v}\left(-x_e(z)\ln x_e(z)+(1-x_e(z))\ln(1-x_e(z))\right)\right.\\ &\qquad\left.-2\left(1-\sum_{e\in\partial v}x_e(z)\right)\ln\left(1-\sum_{e\in\partial v}x_e(z)\right)\right\}.\end{aligned}$$

Hence, we obtained an explicit formula for $\Phi_G(z)=\ln P_G(z)=-U_g(z)\ln z+S_G(z)$ as a function of the $x_e(z)$.

4 A variational formulation

In view of the formula obtained for trees, we define for $\mathbf{x} \in FM(G)$:

$$
\begin{aligned}
U_G^B(\mathbf{x}) &= -\sum_{e \in E} x_e, \\
S_G^B(\mathbf{x}) &= \frac{1}{2} \sum_{v \in V} \left\{ \sum_{e \in \partial v} -x_e \ln x_e + (1-x_e)\ln(1-x_e) \right. \\
&\quad \left. -2\left(1 - \sum_{e \in \partial v} x_e\right) \ln\left(1 - \sum_{e \in \partial v} x_e\right)\right\}, \\
\Phi_G^B(\mathbf{x}, z) &= -U_G^B(\mathbf{x}) \ln z + S_G^B(\mathbf{x}),
\end{aligned}
$$

with the standard convention $0 \ln 0 = 0$. $\Phi_G^B(\mathbf{x}, z)$ is called the Bethe free entropy.

Proposition 9. *Recall that $\mathbf{x}(z) \in \mathbb{R}^E$ is defined by (12). Then we have, for a general graph G:*

$$
\sup_{\mathbf{x} \in FM(G)} \Phi_G^B(\mathbf{x}; z) = \Phi_G^B(\mathbf{x}(z); z).
$$

Moreover, the function $S_G^B(\mathbf{x})$ is non-negative, concave on $FM(G)$ and

$$
\frac{d\Phi_G^B(\mathbf{x}(z); z)}{dz} = -\frac{U_G^B(\mathbf{x}(z))}{z}. \tag{17}
$$

Proof. We first prove the second point. For $k \in \mathbb{N}$, we define $\Delta^k = \{\mathbf{x} \in \mathbb{R}^k,\ x_i \geq 0, \sum_{i=1}^k x_i \leq 1\}$.

Lemma 10. *Let $g : \Delta^k \to \mathbb{R}$ be defined by*

$$
\begin{aligned}
g(\mathbf{x}) &= -\sum_i x_i \ln x_i + \sum_i (1-x_i)\ln(1-x_i) \\
&\quad -2\left(1 - \sum_i x_i\right) \ln\left(1 - \sum_i x_i\right).
\end{aligned}
$$

For $k \geq 1$, g is concave. Moreover, we have

$$
\frac{\partial g}{\partial x_i} = \ln\left(\frac{\left(1 - \sum_j x_j\right)^2}{x_i(1-x_i)}\right).
$$

Proof. From Theorem 20 in [12], we know that the function

$$
\begin{aligned}
h(\mathbf{x}) &= -\sum_i x_i \ln x_i + \sum_i (1-x_i)\ln(1-x_i) \\
&\quad -\left(1 - \sum_i x_i\right) \ln\left(1 - \sum_i x_i\right) \\
&\quad +\left(\sum_i x_i\right) \ln\left(\sum_i x_i\right)
\end{aligned}
$$

is non-negative and concave on Δ^k. We have

$$g(\mathbf{x}) = h(\mathbf{x}) + H\left(\sum_i x_i\right),$$

where $H(p) = -p \ln p - (1-p)\ln(1-p)$ is the entropy of a Bernoulli random variable and is concave in p. □

We now prove the proposition. For $e = (uv) \in E$ and $\mathbf{x} \in \mathring{\Delta}^k$ (the interior of Δ^k), we have

$$\frac{\partial \Phi_G^B(\mathbf{x}; z)}{\partial x_e} = -\ln z + \ln\left(\frac{\left(1 - \sum_{f\in\partial v} x_f\right)\left(1 - \sum_{f\in\partial u} x_f\right)}{x_e(1-x_e)}\right).$$

Hence, we have $\frac{\partial \Phi_G^B(\mathbf{x};z)}{\partial x_e} = 0$ if and only if

$$x_e(1-x_e) = z\left(1 - \sum_{f\in\partial v} x_f\right)\left(1 - \sum_{f\in\partial u} x_f\right). \tag{18}$$

We now show that this equality is valid when evaluated at $\mathbf{x}(z)$. Note that $\sum_{f\in\partial v} x_f(z) = \mathcal{D}_v(\mathbf{y}(z))$, so that we have by (9)

$$\begin{aligned}\left(1 - \sum_{f\in\partial v} x_f(z)\right) &= \left(1 - \frac{\sum_{\overrightarrow{e}\in\partial v} Y_{\overrightarrow{e}}(z)}{1+\sum_{\overrightarrow{e}\in\partial v} y_{\overrightarrow{e}}(z)}\right)\\ &= \left(1 + \sum_{\overrightarrow{e}\in\partial v} y_{\overrightarrow{e}}(z)\right)^{-1}\end{aligned}$$

We have for $e = (uv) \in E$,

$$x_e(z) = \frac{y_{u\to v}(z)}{\frac{z}{y_{v\to u}(z)} + y_{u\to v}(z)},$$

and using the fact that $\mathbf{y}(z) = z\mathcal{R}_G(\mathbf{y}(z))$, we get

$$\begin{aligned}x_e(z) &= \frac{y_{u\to v}(z)}{1+\sum_{w\in\partial v} y_{w\to v}(z)} = y_{u\to v}(z)\left(1 - \sum_{f\in\partial v} x_f(z)\right)\\ 1 - x_e(z) &= \frac{1+\sum_{w\in\partial u\setminus v} y_{w\to u}(z)}{1+\sum_{w\in\partial u} y_{w\to u}(z)} = \frac{z}{y_{u\to v}(z)}\left(1 - \sum_{f\in\partial u} x_f(z)\right),\end{aligned}$$

and we see that (18) is true when evaluated at $x_e(z)$. Hence we proved that $\frac{\partial \Phi_G^B(\mathbf{x}(z);z)}{\partial x_e} = 0$ and the proposition follows. □

5 Gibbs measure on an infinite tree

Let $T = (V, E)$ be an infinite tree with bounded degree. Clearly the definition (2) does not make sense anymore. But we can still define the map $\mathcal{R}_T : (0,\infty)^{\overrightarrow{E}} \to (0,\infty)^{\overrightarrow{E}}$ by $\mathcal{R}_T(\mathbf{a}) = \mathbf{b}$ with

$$b_{u\to v} = \frac{1}{1 + \sum_{w\in\partial u\backslash v} a_{w\to u}},$$

with the convention that the sum over the empty set equals zero. We also denote by $\mathcal{R}_{u\to v} : (0,\infty)^{\partial u\backslash v} \to (0,\infty)$ the local mapping defined by: $b_{u\to v} = \mathcal{R}_{u\to v}(\mathbf{a})$ (note that only the coordinates of $\mathbf{a}$ in $\partial u\backslash v$ are taken as input of $\mathcal{R}_{u\to v}$). Comparisons between vectors are always componentwise.

A crucial result for the monomer-dimer model is that Proposition 5 extends to infinite trees with bounded degree.

Proposition 11. *Let $T = (V, E)$ be an infinite tree with bounded degree. For each $z > 0$, there exists a unique attractive solution in $(0,\infty)^{\overrightarrow{E}}$ to the fixed point equation $\mathbf{y}(z) = z\mathcal{R}_T(\mathbf{y}(z))$, i.e. such that*

$$y_{u\to v}(z) = \frac{z}{1 + \sum_{w\in\partial u\backslash v} y_{w\to u}(z)}. \tag{19}$$

Moreover the map $z \mapsto \mathbf{y}(z)$ is non-decreasing (component-wise) and the map $z \mapsto \frac{\mathbf{y}(z)}{z}$ is non-increasing on $(0,\infty)$.

Proof. First note that any non-negative solution must satisfy $y_{u\to v}(z) \le z$ for all $(uv) \in E$. The compactness of $[0,z]^{\overrightarrow{E}}$ (as a countable product of compact spaces) guarantees the existence of a solution by Schauder fixed point theorem. Alternatively, as in the proof of Proposition 5, we can define the sequence of messages $\mathbf{x}^t(z)$ by (10) so that $\mathbf{x}^{2t}(z) \nearrow \mathbf{x}^-(z)$ and $\mathbf{x}^{2t+1}(z) \searrow \mathbf{x}^+(z)$.

To prove the uniqueness, we follow the approach in [3]. First, we define the change of variable: $h_{u\to v} = -\ln \frac{y_{u\to v}(z)}{z}$ so that (19) becomes:

$$h_{u\to v} = \ln\left(1 + z \sum_{w\in\partial u\backslash v} e^{-h_{w\to u}}\right). \tag{20}$$

We define the function $f : [0,+\infty)^d \mapsto [0,\infty)$ as:

$$f(\mathbf{h}) = \ln\left(1 + z\sum_{i=1}^{k} \frac{1}{1 + z\sum_{j=1}^{k_i} e^{-h^i_j}}\right),$$

where the parameters k, k_i and z are fixed and $d = \sum_{i=1}^{k} k_i$.

Iterating the recursion (20), we can rewrite it using such a function f so that uniqueness would be implied if we show that f is contracting.

For any $\mathbf{h}$ and $\mathbf{h}'$, we apply the mean value theorem to the function $f(\alpha\mathbf{h} + (1-\alpha)\mathbf{h}')$ so that there exists $\alpha \in [0,1]$ such that for $\mathbf{h}_\alpha = \alpha\mathbf{h} + (1-\alpha)\mathbf{h}'$,

$$|f(\mathbf{h}) - f(\mathbf{h}')| = |\nabla f(\mathbf{h}_\alpha)(\mathbf{h} - \mathbf{h}')| \le \|\nabla f(\mathbf{h}_\alpha)\|_{L_1}\|\mathbf{h} - \mathbf{h}'\|_\infty.$$

A simple computation shows that:

$$\|\nabla f(\mathbf{h})\|_{L_1} = \frac{z\sum_{i=1}^{k} \frac{z\sum_{j=1}^{k_i} e^{-h_j^i}}{\left(1+z\sum_{j=1}^{k_i} e^{-h_j^i}\right)^2}}{1+z\sum_{i=1}^{k} \frac{1}{1+z\sum_{j=1}^{k_i} e^{-h_j^i}}}.$$

Let $A_i = \left(1+z\sum_{j=1}^{k_i} e^{-h_j^i}\right)^{-1}$, then we get

$$\|\nabla f(\mathbf{h})\|_{L_1} = \frac{z\sum_{i=1}^{k}(A_i - A_i^2)}{1+z\sum_{i=1}^{k} A_i} = 1 - \frac{1+z\sum_{i=1}^{k} A_i^2}{1+z\sum_{i=1}^{k} A_i}.$$

By taking the partial derivatives, we note that this last expression is maximized when all A_i are equal. Then the solution for the optimal A_i reduces to a quadratic equation with solution in $[0,+\infty)$ equals to $A_i = \frac{\sqrt{1+kz}-1}{kz}$. Substituting for the maximum value, we get for any real vector $\mathbf{h}$,

$$\|\nabla f(\mathbf{h})\|_{L_1} \le 1 - \frac{2}{\sqrt{1+kz}+1}.$$

We now prove that $z \mapsto \frac{\mathbf{x}^t(z)}{z}$ and $z \mapsto \mathbf{x}^t(z)$ are respectively non-increasing and non-decreasing, this implies the last point. We prove it by induction on t: consider $z \le z'$ if $\mathbf{x}^t(z) \le \mathbf{x}^t(z')$ then by (10) we have $\frac{\mathbf{x}^{t+1}(z)}{z} \ge \frac{\mathbf{x}^{t+1}(z')}{z'}$ and if $\frac{\mathbf{x}^t(z)}{z} \ge \frac{\mathbf{x}^t(z')}{z'}$ then again by (10), we have $\mathbf{x}^{t+1}(z) \le \mathbf{x}^{t+1}(z')$. □

We can now compute a Gibbs measure μ_T^z on T by analogy with the finite case: first define $\mathbf{x}(z) \in \mathbb{R}^E$ by (12) and the marginals $\mu_{\partial v}$ and μ_e for each $v \in V$ and $e \in E$ by (13) and (14) respectively. We then define μ_T^z as follows: for any finite vertex-induced subtree $T_f = (V_f, E_f)$ of T, we define:

$$\mu_T^z(b_e, e \in T_f) = \frac{\prod_{v\in V_f} \mu_{\partial v}(\mathbf{b}_{\partial v})}{\prod_{e\in E_f} \mu_e(b_e)}. \tag{21}$$

Theorem 12. *For any $z > 0$ and infinite tree $T = (V, E)$, the measure μ_T^z defined by (21) is the unique distribution over matchings in T satisfying the following property: for any $e = (uv) \in E$ and $k \in \mathbb{N}$, let T_e^k be the subtree of T induced by all vertices at distance k from either u or v, then*

$$\lim_{k\to\infty} \mu_{T_e^k}^z(B_e = 1) = \mu_T^z(B_e = 1).$$

Moreover, we have for $k \ge 0$

$$\mu_{T_e^{2k}}^z(B_e = 1) \le \mu_T^z(B_e = 1) \le \mu_{T_e^{2k+1}}^z(B_e = 1) \tag{22}$$

Proof. The proof follows by checking the compatibility condition of the Kolmogorov extension theorem. The inequality (22) follows by a simple induction on k following the same argument as in the proof of Proposition 5. Taking the limit $k \to \infty$ in (22), we obtain the last claim by invoking Proposition 11. □

Example 13. *The Monomer-Dimer model on the infinite line. In this case, since each vertex has degree* 2*, we see that the solution to (19) must solve:*

$$y(z) = \frac{z}{1+y(z)}.$$

Hence, we have $y(z) = \frac{\sqrt{1+4z}-1}{2}$*, so that*

$$x_e(z) = \frac{y(z)^2}{z+y(z)^2} = \frac{1+2z-\sqrt{1+4z}}{1+4z-\sqrt{1+4z}}.$$

We can for example compute:

$$\mu_{\partial v}\left(\sum_{e\in\partial v} B_e = 0\right) = 1 - 2x_e(z) = \frac{1}{\sqrt{1+4z}}.$$

We can check that using (9):

$$\mathcal{D}_v(y(z)) = \frac{2y(z)}{1+2y(z)} = 1 - \frac{1}{\sqrt{1+4z}}.$$

6 Lifts of graph: definitions

If G is a graph and $v \in V(G)$, the 1-neighbourhood of v is the subgraph consisting of all edges incident upon v. A graph homomorphism $\pi : G' \to G$ is a covering map if for each $v' \in V(G')$, π gives a bijection of the edges of the 1-neighbourhood of v' with those of $v = \pi(v')$. G' is a cover or a lift of G.

Given a graph G with a distinguished vertex $v \in V$, we construct the (infinite) rooted tree $(T(G), v)$ of non-backtracking walks at v as follows: its vertices correspond to the finite non-backtracking walks in G starting in v, and we connect two walks if one of them is a one-step extension of the other. With a slight abuse of notation, we denote by v the root of the tree of non-backtracking walks started at v. Note that also we constructed $T(G)$ from a particular vertex v, this choice is irrelevant. It is easy to see that $T(G)$ is a cover of G, indeed it is the (unique up to isomorphism) cover of G that is also a cover of every other cover of G. $T(G)$ is called the universal cover of G.

Definition 14. *Let G be a graph with no loop. Then H is a 2-lift of G if $V(H) = V(G) \times \{0,1\}$ and for every $(u,v) \in E(G)$, exactly one of the following two pairs are edges of H: $((u,0),(v,0))$ and $((u,1),(v,1)) \in E(H)$ or $((u,0),(v,1))$ and $((u,1),(v,0)) \in E(H)$. If $(u,v) \notin E(G)$, then none of $((u,0),(v,0))$,$((u,1),(v,1))$, $((u,0),(v,1))$ and $((u,1),(v,0))$ are edges in H.*

Definition 15. *Let $G = (V,E)$ be a fixed connected multigraph with no loop. A n-lift of G is a graph on vertex set $V_1 \cup V_2 \cup \cdots \cup V_{v(G)}$, where each V_i is a set of n vertices and these sets are pairwise disjoint, obtained by placing a perfect matching between V_i and V_j for each edge $e = (ij)$ of G.*

The crucial property first proved by Csikvári [6] is:

Proposition 16. *Let G be a bipartite graph and H be a 2-lift of G. Then $P_G(z)^2 \geq P_H(z)$ for $z > 0$.*

Proof. Note that $G \cup G$ is a particular 2-lift of G with $P_{G\cup G}(z) = P_G(z)^2$. To prove the first statement of the proposition, we need to show that for any 2-lift H of G, we have: $m_k(G\cup G) \geq m_k(H)$. Consider the projection of a matching of a 2-lift of G to G. It will consist of disjoint union of cycles of even lengths (since G is bipartite), paths and double-edges when two edges project to the same edge. For such a projection $R = R_1 \cup R_2 \subset E$ where R_2 is the set of double edges. Now for such a projection, we count the number of possible matchings in $G \cup G$: $n_R(G \cup G) = 2^{k(R)}$, where $k(R)$ is the number of connected components of R_1. The number of possible matchings in H is $n_R(H) \leq 2^{k(R)}$ since in each component if the inverse image of one edge is fixed then the inverse images of all other edges is also determined. There is no equality as in general not every cycle can be obtained as a projection of a matching of a 2-lift. For example, if one considers a 8-cycle as a 2-lift of a 4-cycle, then no matching will project on the whole 4-cycle. Hence we proved that $m_k(G\cup G) \geq m_k(H)$ so that $P_G(z)^2 \geq P_H(z)$ for $z > 0$. □

7 Rooted unlabeled graphs

A *rooted graph* (G, o) is a graph $G = (V, E)$ together with a distinguished vertex $o \in V$, called the *root*. We let $\mathcal{G}_\star$ denote the set of all locally finite connected rooted graphs considered up to *rooted isomorphism*, i.e. $(G, o) \equiv (G', o')$ if there exists a bijection $\gamma\colon V \to V'$ that preserves roots ($\gamma(o) = o'$) and adjacency ($\{i, j\} \in E \iff \{\gamma(i), \gamma(j)\} \in E'$). We write $[G, o]_h$ for the (finite) rooted subgraph induced by the vertices lying at graph-distance at most $h \in \mathbb{N}$ from o. The distance

$$\textsc{dist}\left((G, o), (G', o')\right) := \frac{1}{1+r} \quad \text{where} \quad r = \sup\left\{h \in \mathbb{N}\colon [G, o]_h \equiv [G', o']_h\right\},$$

turns $\mathcal{G}_\star$ into a complete separable metric space, see [1]. We will also need edge-rooted graphs and define $\mathcal{G}_{\star\star}$: the space of locally finite connected graphs with a distinguished oriented edge, taken up to the natural isomorphism relation and equipped with the natural distance, which turns it into a complete separable metric space.

With a slight abuse of notation, (G, o) will denote an equivalence class of rooted graph also called unlabeled rooted graph in graph theory terminology. Note that if two rooted graphs are isomorphic, then their rooted trees of non-backtracking walks are also isomorphic. It thus makes sense to define $(T(G), o)$ for elements $(G, o) \in \mathcal{G}_\star$.

Proposition 17. *For any graph $G = (V, E)$, there exists a graph sequence $\{G_n\}_{n\in\mathbb{N}}$ such that $G_0 = G$, G_n is a 2-lift of G_{n-1} for $n \geq 1$. Hence G_n is a 2^n-lift of G and we denote by $\pi_n : G_n \to G$ the corresponding covering. For any $v \in V$, we have:*

$$\sup_{u\in\pi_n^{-1}(v)} \textsc{dist}\left((G_n, u), (T(G), v)\right) \to 0,$$

in particular for any $v_n \in \pi_n^{-1}(v)$, we have $(G_n, v_n) \to (T(G), v)$ in $\mathcal{G}_\star$.

Proof. The proof follows from an argument of Nathan Linial [9], see also [6].

A random 2-lift H of a base graph G is the random graph obtained by choosing between the two pairs of edges $((u,0),(v,0))$ and $((u,1),(v,1)) \in E(H)$ or $((u,0),(v,1))$ and $((u,1),(v,0)) \in E(H)$ with probability 1/2 and each choice being made independently.

Let G be a graph with girth γ and let k be the number of cycles in G with size γ. Let X be the number of γ-cycles in H a random 2-lift of G. The girth of H must be at least γ and a γ-cycle in H must be a lift of a γ-cycle in G. A γ-cycle in G yields: a 2γ-cycle in H with probability 1/2; or two γ-cycles in H with probability 1/2. Hence we have $\mathbb{E}[X] = k$. But $G \cup G$ (the trivial lift) has $2k$ γ-cycles. Hence there exists a 2-lift with strictly less than k γ-cycles. By iterating this step, we see that there exists a sequence $\{G_n\}$ of 2-lifts such that for any γ, there exists a $n(\gamma)$ such that for $j \geq n(\gamma)$, the graph G_j has no cycle of length at most γ. This implies that for any $v \in V$ and $v_j \in \pi_j^{-1}(v)$, we have $\text{DIST}\left((G_j, v_j), (T(G), v)\right) \leq \frac{2}{\gamma}$ and the proposition follows. □

8 Thermodynamic limit (for lifts)

Proposition 18. *Let T be a tree with bounded degree. If $(G_n, e_n) \to (T, e)$ in $\mathcal{G}_{\star\star}$, then for any $z > 0$,*

$$\lim_{n\to\infty} \mu^z_{G_n}(B_{e_n} = 1) = \mu^z_T(B_e = 1)$$

Proof. By assumption, for any radius $h \in \mathbb{N}$, there exists n_h such that for all $n \geq n_h$, we have $[G_n, e_n]_h \equiv [T, e]_h$. Clearly for any graph $G = (V, E)$ and $e \in E$, $\mu^z_G(B_e = 1)$ depends only on the isomorphism class of the edge-rooted graph (G, e). The claim then follows from the Markov property (Lemma 4) and Theorem 12. □

Applying this result together with Proposition 17, we obtain, for any $v \in V$,

$$\lim_{n\to\infty} \sup_{u\in\pi_n^{-1}(v)} \left(\sum_{e\in\partial u} \mu^z_{G_n}(B_e = 1) - \sum_{e\in\partial v} \mu^z_{T(G)}(B_e = 1) \right) = 0. \tag{23}$$

We now show that thanks to the particular structure of $T(G)$, we are able to extend this result to show the convergence of $\frac{1}{|V_n|} U_{G_n}(z)$, $\frac{1}{|V_n|} S_{G_n}(z)$, or $\frac{1}{|V_n|} \ln P_{G_n}(z)$.

The crucial observation is the following. Since the local recursions are the same for both $\mathcal{R}_{T(G)}$ and $\mathcal{R}_G$ and since there is a unique fixed point for both $z\mathcal{R}_{T(G)}$ and $z\mathcal{R}_G$, the proposition below follows:

Proposition 19. *Let G be a finite graph and $T(G)$ be its universal cover and associated cover $\pi : T(G) \to G$. By Propositions 5 and 11, we can define:*

$$\tilde{\mathbf{y}}(z) = z\mathcal{R}_{T(G)}(\tilde{\mathbf{y}}(z)) \text{ , and, } \mathbf{y}(z) = z\mathcal{R}_G(\mathbf{y}(z)).$$

We have $\pi(\tilde{\mathbf{y}}(z)) = \mathbf{y}(z)$, i.e. $\tilde{y}_{\overrightarrow{e}}(z) = y_{\pi(\overrightarrow{e})}(z)$.

We are now ready to prove

Theorem 20. *Let G be a finite graph and $(G_n)_{n\geq 1}$ be the sequence defined in Proposition 17 Then we have as $n \to \infty$, for $z > 0$,*

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{|V_n|} \ln P_{G_n}(z) &= \frac{1}{v(G)} \Phi_G^B(\mathbf{x}(z), z), & (24)\\
\lim_{n\to\infty} \frac{1}{|V_n|} U_{G_n}(z) &= \frac{1}{v(G)} U_G^B(\mathbf{x}(z)), & (25)\\
\lim_{n\to\infty} \frac{\nu(G_n)}{|V_n|} &= \frac{\nu^*(G)}{v(G)}, & (26)\\
\lim_{n\to\infty} \frac{1}{|V_n|} S_{G_n}(z) &= \frac{1}{v(G)} S_G^B(\mathbf{x}(z)). & (27)
\end{aligned}$$

Proof. We write:

$$\begin{aligned}
\frac{1}{|V_n|} U_{G_n}(z) &= \frac{-1}{2|V_n|} \sum_{v\in V_n} \sum_{e\in\partial v} \mu_{G_n}^z(B_e = 1)\\
&= \frac{-1}{2|V|} \sum_{v\in V} \frac{1}{2^n} \sum_{u\in\pi_n^{-1}(v)} \sum_{e\in\partial u} \mu_{G_n}^z(B_e = 1).
\end{aligned}$$

Recall that $|\pi_n^{-1}(v)| = 2^n$ for each $v \in V$ so that by (23), we have

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{|V_n|} U_{G_n}(z) &= \frac{-1}{2|V|} \sum_{v\in V} \sum_{e\in\partial v} \mu_{T(G)}^z(B_e = 1)\\
&= \frac{-1}{2|V|} \sum_{v\in V} \sum_{e\in\partial v} x_e(z),
\end{aligned}$$

where the last equality follows from Proposition 19 and $\mathbf{x}(z)$ is defined by (12). Hence we proved (25).

We now prove (26). By Lemma 3, we have

$$\frac{-U_G^B(\mathbf{x}(z))}{v(G)} \leq \liminf_{n\to\infty} \frac{\nu(G_n)}{|V_n|} \leq \limsup_{n\to\infty} \frac{\nu(G_n)}{|V_n|} \leq \frac{-U_G^B(\mathbf{x}(z))}{v(G)} + \frac{|E|\ln 2}{|V|\ln z}$$

Since $\Phi_G^B(\mathbf{x}(z), z) = -U_G^B(\mathbf{x}(z))\ln(z) + S_G^B(\mathbf{x}(z))$ and $S_G^B(\mathbf{x})$ is bounded, we have $\lim_{z\to\infty} \frac{\Phi_G^B(\mathbf{x}(z),z)}{\ln z} = \lim_{z\to\infty} -U_G^B(\mathbf{x}(z))$. By Proposition 9, we have

$$\frac{\Phi_G^B(\mathbf{x}(z), z)}{\ln z} = \sup_{\mathbf{x}\in FM(G)} \sum_{e\in E} x_e + \frac{S_G^B(\mathbf{x})}{\ln z}$$

so that $\lim_{z\to\infty} \frac{\Phi_G^B(\mathbf{x}(z),z)}{\ln z} = \nu^*(G)$ and (26) follows.

We now prove (24). Recall that $\Phi_G'(z) = -\frac{U_G(z)}{z}$ so that the convergence of $\frac{1}{|V_n|}\ln P_{G_n}(z)$ follows from (25) and Lebesgue dominated convergence theorem. We first need to check that the derivative with respect to z of the right-hand term in (24) is $\frac{U_G^B(\mathbf{x}(z))}{z}$ and this is exaclty what we showed in Proposition 9, see (17). In order to conclude the proof of (24), we will show that

$$\lim_{n\to\infty} \lim_{z\to\infty} \frac{1}{|V_n|} \frac{\ln P_{G_n}(z)}{\ln z} = \lim_{z\to\infty} \frac{1}{v(G)} \frac{\Phi_G^B(\mathbf{x}(z), z)}{\ln z} = \frac{\nu^*(G)}{v(G)}.$$

For the left-hand term, we have

$$\frac{1}{|V_n|}\frac{\ln P_{G_n}(z)}{\ln z} \geq -U_{G_n}(z),$$

and since the number of matching is upper bounded by $2^{|E_n|}$, we have

$$\frac{1}{|V_n|}\frac{\ln P_{G_n}(z)}{\ln z} \leq -U_{G_n}(z) + \frac{|E_n|\ln 2}{|V_n|\ln z}.$$

Hence, taking the limit $z \to \infty$, we have

$$\lim_{z\to\infty}\frac{1}{|V_n|}\frac{\ln P_{G_n}(z)}{\ln z} = \frac{\nu(G_n)}{|V_n|}.$$

which together with (26) concludes the proof of (24). Finally (27) follows from (25) and (24). □

9 Application: a lower bound for bipartite graphs

In the special case where G is a bipartite graph, we will prove the following lower bound:

Theorem 21. *For any finite bipartite graph G, we have for $z > 0$,*

$$\ln P_G(z) \geq \max_{\mathbf{x}\in M(G)}\left\{\left(\sum_e x_e\right)\ln z + S_G^B(\mathbf{x})\right\}. \tag{28}$$

Proof. The crucial property first proved by Csikvári [6] is:

Proposition 22. *Let G be a bipartite graph and H be a 2-lift of G. Then $P_G(z)^2 \geq P_H(z)$ for $z > 0$.*

Proof. Note that $G \cup G$ is a particular 2-lift of G with $P_{G\cup G}(z) = P_G(z)^2$. To prove the first statement of the proposition, we need to show that for any 2-lift H of G, we have: $m_k(G\cup G) \geq m_k(H)$. Consider the projection of a matching of a 2-lift of G to G. It will consist of disjoint union of cycles of even lengths (since G is bipartite), paths and double-edges when two edges project to the same edge. For such a projection $R = R_1 \cup R_2 \subset E$ where R_2 is the set of double edges. Now for such a projection, we count the number of possible matchings in $G \cup G$: $n_R(G\cup G) = 2^{k(R)}$, where $k(R)$ is the number of connected components of R_1. The number of possible matchings in H is $n_R(H) \leq 2^{k(R)}$ since in each component if the inverse image of one edge is fixed then the inverse images of all other edges is also determined. There is no equality as in general not every cycle can be obtained as a projection of a matching of a 2-lift. For example, if one considers a 8-cycle as a 2-lift of a 4-cycle, then no matching will project on the whole 4-cycle. Hence we proved that $m_k(G\cup G) \geq m_k(H)$ so that $P_G(z)^2 \geq P_H(z)$ for $z > 0$. □

Hence, if we consider the sequence of 2-lifts constructed in Proposition 17, we have the sequence $\{\frac{1}{|V_n|}\Phi_{G_n}(z)\}_{n\in\mathbb{N}}$ is non-increasing in n and converges to $\frac{1}{v(G)}\Phi_G^B(\mathbf{x}(z), z)$ by Theorem 20. The claim follows from Proposition 9 and the fact that $FM(G) = M(G)$ for a bipartite graph. □

10 The framework of local weak convergence

This section gives a brief account of the framework of local weak convergence. For more details, we refer to the surveys [2, 1].

Let $\mathcal{P}(\mathcal{G}_\star)$ denote the set of Borel probability measures on $\mathcal{G}_\star$, equipped with the usual topology of weak convergence (see e.g. [4]). Given a finite graph $G = (V, E)$, we construct a random element of $\mathcal{G}_\star$ by choosing uniformly at random a vertex $o \in V$ to be the root, and restricting G to the connected component of o. The resulting law is denoted by $\mathcal{U}(G)$. If $\{G_n\}_{n\geq 1}$ is a sequence of finite graphs such that $\{\mathcal{U}(G_n)\}_{n\geq 1}$ admits a weak limit $\mathcal{L} \in \mathcal{P}(\mathcal{G}_\star)$, we call $\mathcal{L}$ the *local weak limit* of $\{G_n\}_{n\geq 1}$. If (G, o) denotes a random element of $\mathcal{G}_\star$ with law $\mathcal{L}$, we shall use the following slightly abusive notation : $G_n \rightsquigarrow (G, o)$ and for $f : \mathcal{G}_\star \to \mathbb{R}$:

$$\mathbb{E}_{(G,o)}\left[f(G,o)\right] = \int_{\mathcal{G}_\star} f(G,o) d\mathcal{L}(G,o).$$

Unimodularity. Recall that $\mathcal{G}_{\star\star}$ denotes the space of locally finite connected graphs with a distinguished oriented edge, taken up to the natural isomorphism relation and equipped with the natural distance, which turns it into a complete separable metric space. With $f : \mathcal{G}_{\star\star} \to \mathbb{R}$, we associate a function $\partial f : \mathcal{G}_\star \to \mathbb{R}$, defined by:

$$\partial f(G,o) = \sum_{i \in \partial o} f(G,o,i),$$

and also the reversal $f^* : \mathcal{G}_{\star\star} \to \mathbb{R}$ of f defined by:

$$f^*(G,o,i) = f(G,i,o).$$

It is shown in [1] that any (G, o) with law $\mathcal{L}$ arising as the local weak limit of some sequence of finite graphs satisfies

$$\mathbb{E}_{(G,o)}\left[\partial f(G,o)\right] = \mathbb{E}_{(G,o)}\left[\partial f^*(G,o)\right] \tag{29}$$

for any Borel $f : \mathcal{G}_{\star\star} \to [0, \infty)$. A measure $\mathcal{L} \in \mathcal{P}(\mathcal{G}_\star)$ satisfying this invariance is called *unimodular*, and the set of all unimodular probability measures on $\mathcal{G}_\star$ is denoted by $\mathcal{P}_u(\mathcal{G}_\star)$. Note that (29) can be expanded to:

$$\int_{\mathcal{G}_\star} \sum_{i \in \partial o} f(G,o,i)\, d\mathcal{L}(G,o) = \int_{\mathcal{G}_\star} \sum_{i \in \partial o} f(G,i,o)\, d\mathcal{L}(G,o).$$

11 Application to matchings

...
We define

$$\Phi_o(\mathbf{x}) = \frac{\ln z}{2} \sum_{e \in \partial o} x_e + \frac{1}{2} \sum_{e \in \partial o} \left(-x_e \ln x_e + (1 - x_e) \ln(1 - x_e)\right) - \left(1 - \sum_{e \in \partial o} x_e\right) \ln \left(1 - \sum_{e \in \partial o} x_e\right).$$

Lemma 23. *We have*

$$\begin{aligned}\Phi_o(\mathbf{x}(z)) &= \left(1-\frac{|\partial o|}{2}\right)\ln\left(1+\sum_{\vec{e}\in\partial o} y_{\vec{e}}(z)\right)+\frac{1}{2}\sum_{\vec{e}\in\partial o}\ln\left(1+\sum_{\vec{f}\in\partial o\setminus\vec{e}} y_{\vec{f}}(z)\right)\\ &= \ln\left(1+\sum_{\vec{e}\in\partial o} y_{\vec{e}}(z)\right)+\frac{1}{2}\sum_{\vec{e}\in\partial o}\ln\left(1-\frac{y_{\vec{e}}(z)}{1+\sum_{\vec{f}\in\partial o} y_{\vec{f}}(z)}\right).\end{aligned}$$

References

[1] D. Aldous and R. Lyons. Processes on unimodular random networks. *Electronic Journal of Probability*, 12:1454–1508, 2007.

[2] D. Aldous and J. M. Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 1–72. Springer, Berlin, 2004.

[3] M. Bayati, D. Gamarnik, D. Katz, C. Nair, and P. Tetali. Simple deterministic approximation algorithms for counting matchings. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, page 127. ACM, 2007.

[4] P. Billingsley. *Convergence of probability measures.* John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[5] C. Bordenave, M. Lelarge, and J. Salez. Matchings on infinite graphs. *Probability Theory and Related Fields*, pages 1–26, 2012.

[6] P. Csikvári. Lower matching conjecture, and a new proof of Schrijver's and Gurvits's theorems. *arXiv preprint arXiv:1406.0766*, 2014.

[7] M. Lelarge. Loopy annealing belief propagation for vertex cover and matching: convergence, lp relaxation, correctness and Bethe approximation. *arXiv preprint arXiv:1401.7923*, 2014.

[8] M. Lelarge. Counting matchings in irregular bipartite graphs and random lifts. *arXiv preprint arXiv:1507.04739*, 2015.

[9] N. Linial. Lifts of graphs (slides).

[10] J. Salez. Weighted enumeration of spanning subgraphs in locally tree-like graphs. *Random Structures & Algorithms*, 2012.

[11] A. Schrijver. *Combinatorial optimization. Polyhedra and efficiency.*, volume 24 of *Algorithms and Combinatorics.* Springer-Verlag, Berlin, 2003.

[12] P. O. Vontobel. The Bethe permanent of a nonnegative matrix. *IEEE Trans. Inform. Theory*, 59(3):1866–1901, 2013.

[13] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.