Witness trees in the Moser-Tardos algorithmic Lovász Local Lemma and Penrose trees in the hard core lattice gas

Aldo Procacci[1] (Joint work with Rogério G. Alves[2])

[1] UFMG Belo Horizonte MG Brazil

[2] UFOP Ouro Preto MG Brazil

work available at http://arxiv.org/abs/1306.3511

The Lovász Local Lemma

The **Local Lemma (LLL), (Erdös Lovász in 1975)** is a tool in the framework of the *probabilistic methods in combinatorics* to prove the existence of combinatorial objects with certain desirable properties (e.g. a proper coloring of the edges of a graph) by showing that these objects have a positive probability to occur in some probability space.

Stating the LLL

Take a finite family $\{A_x\}_{x \in X}$, of (bad) events in some probability space, having probabilities $Prob(A_x) = p_x$. Let $\bar{A}_x$ be the complement event of A_x, so that $\cap_{x \in X} \bar{A}_x$ is the event that none of the events A_x occur (the good event).

When the event $\cap_{x \in X} \bar{A}_x$ has a positive probability to occur (and hence it exists)?

- If events A_x are disjoint (the worst case) then
$P(\cap_{x \in X} \bar{A}_x) = 1 - \sum_{x \in X} p_x$,
hence $\sum_{x \in X} p_x < 1 \implies P(\cap_{x \in X} \bar{A}_x) > 0$.

- On the other hand, if the events $\{A_x\}_{x \in X}$ are independent (the best case), then $P(\cap_{x \in X} \bar{A}_x) = \prod_{x \in X}(1 - p_x)$,
hence $p_x < 1$ for all $x \implies P(\cap_{x \in X} \bar{A}_x) > 0$.

The Lovász local lemma (LLL) deals with the intermediate cases. This lemma ensures that $P(\cap_{x\in X}\bar{A}_x) > 0$ under relatively mild conditions on the $\{p_x\}_{x\in X}$ when there is strong dependence only among some subsets of the $\{A_x\}_{x\in X}$, while most of these events are independent.

Definition (Dependency Graph):. Given a family of events $\{A_x\}_{x\in X}$ on some probability space, a graph G with vertex set $V(G) = X$ is a *dependency graph* for the events $\{A_x\}_{x\in X}$ if, for each $x \in X$, A_x is independent of all the events in the σ-algebra generated by $\{A_y : y \in X \backslash \Gamma^*_G(x)\}$, where $\Gamma_G(x)$ denotes the vertices of G adjacent to x and $\Gamma^*_G(x) = \Gamma_G(x) \cup \{x\}$.

Theorem 1 (Lovász Local Lemma) *Let* $G = (X, E)$ *be a dependence graph for the collection of events* $\{A_x\}_{x\in X}$ *each with probability* $Prob(A_x) = p_x$ *and let* $\boldsymbol{\mu} = \{\mu_x\}_{x\in X}$ *be real numbers in* $[0, +\infty)$*. If, for each* $x \in X$*,*

$$p_x \le \frac{\mu_x}{\prod_{y\in\Gamma^*_G(y)}(1+\mu_y)}$$

then

$$Prob(\bigcap_{x\in X} \bar{A}_x) > 0$$

Let us now (apparently) change subject and talk about the hard core lattice gas on the graph $G = (X, E)$

The Hard-core gas on a graph $G = (X, E)$

- Each vertex $x \in X$ can be occupied by at most one particle or can be left empty.
- the particle occupying the vertex $x \in X$ carries an activity $w_x \in \mathbb{C}$ ($\boldsymbol{w} = \{w_x\}_{x \in X}$ is the set of all activities).
- If a vertex $x \in X$ is occupied, then neighbor vertices of x (i.e. those in $\Gamma_G(x)$) are empty
- The partition function of this gas is defines as

$$\Xi_X(\boldsymbol{w}) = \sum_{\substack{Y \subset X \\ Y \text{ independent in } G}} \prod_{y \in Y} w_y \qquad (1.1)$$

from which one gets the "pressure" and the one-point correlation function

$$P(\boldsymbol{w}) = \frac{1}{|X|} \log \Xi_X(\boldsymbol{w}) \ ; \ \Psi_{x_0}(\boldsymbol{w}) = -w_{x_0} \frac{\partial}{\partial w_{x_0}} \log \Xi_X(\boldsymbol{w}) \qquad (1.2)$$

Known fact: $\log \Xi_X(\boldsymbol{w})$ (hence $P(\boldsymbol{w})$ and $\Psi_{x_0}(\boldsymbol{w})$), can be written as formal series, known as *cluster expansion*. E.g., the one-point correlation is written as

$$\Psi_{x_0}(\boldsymbol{w}) = -\sum_{n=0}^{\infty}\frac{1}{n!}\sum_{(x_1,\dots,x_n)\in X^n}\sum_{\substack{g\in G_n^0\\ g\subset g(x_0,x_1,\dots,x_n)}}(-1)^{|E_g|}\prod_{i=0}^{n} w_{x_i} \tag{1.3}$$

where
- $G_n^0 =$ set of all connected graphs with vertex set $\mathrm{I}_n^0 \doteq \{0,1,\dots,n\}$
- If $(x_0,\dots,x_n)\in X^{n+1}$ then $g(x_0,\dots,x_n)$ is the graph with vertex set I_n^0 which has the edge $\{i,j\}$ if and only if $x_i \not\sim x_j$ (i.e. **if either** $\{x_i,x_j\}\in E$ **or** $x_i = x_j$).

The equation (1.3) makes sense only for those $\boldsymbol{w}\in\mathbb{C}^{|X|}$ such that the formal series in the r.h.s. of (1.3) converge absolutely.

Another known fact: Alternating sign property:

$$\sum_{\substack{g\in G_n^0\\ g\subset g(x_0,x_1,\dots,x_n)}} (-1)^{|E_g|} = (-1)^n \Big| \sum_{\substack{g\in G_n^0\\ g\subset g(x_0,x_1,\dots,x_n)}} (-1)^{|E_g|} \Big|$$

So for $\boldsymbol{\rho} = \{\rho_x\}_{x\in X}$ with $\rho_x \in (0,\infty)$, $\Psi_{x_0}(-\boldsymbol{\rho})$ is

$$\Psi_{x_0}(-\boldsymbol{\rho}) = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{(x_1,\dots,x_n)\in X^n} \Big| \sum_{\substack{g\in G_n^0\\ g\subset g(x_0,x_1,\dots,x_n)}} (-1)^{|E_g|} \Big| \prod_{i=0}^{n} \rho_{x_i} \tag{1.4}$$

I.e. a positive term series. Let

$$\mathcal{R}(G) = \{\boldsymbol{\rho} \in [0,\infty)^{|X|} :\ \Psi_{x_0}(-\boldsymbol{\rho}) < +\infty\}$$

Then, **if** $\rho \in \mathcal{R}(G)$, **then** $\Psi_{x_0}(\boldsymbol{w})$ is analytic for $\boldsymbol{w} = \{w_x\}_{x\in X}$ in the poly-disk $|\boldsymbol{w}| \le \boldsymbol{\rho}$. The set $\mathcal{R}(G)$ is called *convergence region* of the cluster expansion.

Let's give two convergence criteria for $\Psi_{x_0}(-\boldsymbol{\rho})$:

Theorem 2 ((Dobrushin 1996) (Fernández & P.2007))
Give the hard core lattice gas on a graph $G = (X, E)$*, let* $\boldsymbol{\mu} = (\mu_x)_{x \in X}$ *be a family of non negative numbers in* $[0, +\infty)$*. If* $\boldsymbol{\rho} = (\rho_x)_{x \in X}$ *is such that, for all* $x \in X$

$$\rho_x \leq \frac{\mu_x}{\prod_{y \in \Gamma^*_G(y)} (1 + \mu_y)} \quad ; \quad \rho_x \leq \frac{\mu_x}{\sum_{\substack{R \subseteq \Gamma^*_G(x) \\ R \text{ indep in } G}} \prod_{x \in R} \mu_x}$$

then $\boldsymbol{\rho} \in \mathcal{R}(G)$ *and* $\Psi_x(-\boldsymbol{\rho}) \leq \mu_x$

Fernández & P. improves Dobrushin, since

$$\prod_{y \in \Gamma^*_G(x)} (1 + \mu_y) = \sum_{R \subseteq \Gamma^*_G(x)} \prod_{x \in R} \mu_x \geq \sum_{\substack{R \subseteq \Gamma^*_G(x) \\ R \text{ indep in } G}} \prod_{x \in R} \mu_x$$

Connection between LLL and the hard-core gas
Scott and Sokal (2005) pointed out a beautiful connection between LLL and the the hard core gas.

Theorem 3 (Scott-Sokal [via Shearer 1985]) *Let G be a dependence graph for the family of events $\{A_x\}_{x\in X}$ with probability p_x. Let $\Xi_X(\boldsymbol{w})$ be the partition function of the hard core gas on G and let $\mathcal{R}(G)$ the convergence region of the cluster expansion. If $\boldsymbol{p} = \{p_x\}_{x\in X} \in \mathcal{R}(G)$, then,*

$$Prob\Big(\bigcap_{x\in X} \bar{A}_x\Big) \ \geq\ \Xi_X(-\boldsymbol{p}) \ >\ 0\,. \tag{1.5}$$

Furthermore these bounds are the best possible, i.e. if $\boldsymbol{p} \notin \mathcal{R}(G)$, then there is a family of events $\{B_x\}_{x\in X}$ with probabilities $Prob(B_x) = p_x$ and dependency graph G, such that $Prob(\bigcap_{x\in X} \bar{B}_x) = 0$.

By merging Theorem 2 (Dobrushin) into Theorem 3 one obtains the usual LLL, i.e. Theorem 1.
By using Theorem 2 (Fernández & P.) we get

Theorem 4 (Bissacot et al. (2011)) *Under the same hypothesis of Theorem 1, if, for each $x \in X$,*

$$p_x \ \leq \frac{\mu_x}{\sum\limits_{\substack{R \subseteq \Gamma^*_G(x) \\ R \text{ indep in } G}} \prod_{x \in R} \mu_x} \tag{1.6}$$

Then

$$Prob\Big(\bigcap_{x \in X} \bar{A}_x\Big) \ > \ 0$$

This **improved version of LLL** has been already used to obtain improved bounds on various graph coloring problems (Ndreca and P. EJC 2013).

The Algorithmic Moser-Tardos version of the Lovász Local Lemma

Criticism to LLL: its inherently non-constructive character: LLL says the good event exists, but is there an algorithm capable to find it, possibly in a polynomial time?

Efforts to devise an algorithmic version of the LLL culminate in a recent breakthrough paper by Moser and Tardos (2009), who gave a fully algorithmic version of LLL if the events are restricted to a class which however covers basically all known applications of LLL.

Moser Tardos setting

- Suppose to have a finite collection of mutually independent random variables Π.

- Let $\boldsymbol{A} = \{A_x\}_{x \in X}$ be a finite family of events depending onthese variables Π, each A_x with probability $Prob(A_x)$ such that:
each A_x depends only on some subset $vbl(A_x) \subset \Pi$ of the whole set of random variables of the family Π.

- The dependence graph of the family $\boldsymbol{A}$ is the graph $G = (X, E)$ with vertex set X and edge set E constituted by the pairs $\{x, x'\} \subset X$ s. t. $vbl(A_x) \cap vbl(A_{x'}) \neq \emptyset$.

Obs: a lot of application of LLL are in this setting

Theorem 5 (Moser Tardos) *Let* $\boldsymbol{A} = \{A_x\}_X$ *be a finite set of events determined by variables* $\nu \in \Pi$ *as above, each with probability* p_x *and with dependency graph* G*. Let* $\boldsymbol{\mu} = (\mu_x)_{x \in X}$ *be a a sequence of real numbers in* $[0, +\infty)$*. If, for each* $x \in X$*,*

$$p_x \leq \frac{\mu_x}{\prod_{y \in \Gamma^*_G(y)} (1 + \mu_y)} \tag{1.7}$$

then there exists an algorithm which finds an assignment of values to the variables $\nu \in \Pi$ *such that none of the events in* $\boldsymbol{A}$ *occurs.*

Moreover the expected total number of steps of the algorithm is at most $\sum_{x \in X} \mu_x$*.*

MT Algorithm is as simple as it can be:

- step 0 : choose a random evaluation of the variables $\nu \in \Pi$

- step i: if some $A \in \boldsymbol{A}$ occurs, pick one, say A_x and take a new evaluation (resampling) only of its variables $vbl(A_x)$,keeping unchanged all the other variables in Π.

The algorithm stops when we reach an evaluation of the variables $\nu \in \Pi$ such that none of the events in the family $\boldsymbol{A}$ occurs.

Questions:

- is there a connection between the algorithmic LLL and the hard core gas *a la* Scott-Sokal?

- and if yes, can this connection leads to an improvement of the ALLL similarly to the non-constructive case?

Remark. the scheme proposed by Moser and Tardos to prove their Theorem 5, based on the concept of witness trees has nothing to do with the proof (by induction) of the non-algorithmic Lovász Local Lemma.

Strong indications that the connection exists

- Kolipaka and Szegedy (2011) relate the Moser Tardos algorithm to the set of Shearer conditions, however without giving any explicit improvement on Theorem 5

- Pegden (2013) modified the branching process argument to adapt it to the Bissacot et al. condition of Theorem 4.

Theorem 6 *Under the same hypothesis of Theorem 5, let $\mathcal{R}(G)$ be the convergence region of the CE of the hard-core gas on the dependency graph $G = (X, E)$.*

If $\boldsymbol{p} = \{p_x\}_{x \in X} \in \mathcal{R}(G)$, then the MT algorithm finds an assignment of values to the variables Π such that none of the events in $\boldsymbol{A}$ occurs by resampling each event $A_x \in \boldsymbol{A}$ in an expected number os steps N_x s. t.

$$N_x \leq \Psi_{x_0}(-\boldsymbol{p}) \tag{1.8}$$

where $\Psi_{x_0}(-\boldsymbol{p})$ is the one-point correlation function defined above and the expected total number of steps T of the algorithm is at most [$P(\boldsymbol{w})$ = pressure of the lattice gas]

$$T \leq \sum_{x \in X} \Psi_{x_0}(-\boldsymbol{p})$$

Remark 1. Theorem 6 above together with Theorem 2 (with FP) immediately yields for free the following corollary

Corollary 7 (Pegden) *Under the hypothesis of Theorem 6, if* $\boldsymbol{\mu} = \{\mu_x\}_{x \in X}$ *is a sequence of real numbers in* $[0, +\infty)$ *such that, for each* $x \in X$

$$p_x \leq \frac{\mu_x}{\sum\limits_{\substack{R \subseteq \Gamma^*_G(x) \\ R \text{ indep in } G}} \prod_{y \in R} \mu_y}$$

then the randomized algorithm resamples an event $A_x \in \boldsymbol{A}$, *at most an expected* μ_x *times before it finds such an evaluation. Thus the expected total number of resampling steps is at most* $\sum_{x \in X} \mu_x$.

PART 2: Prove of Theorem 6

Witness trees

[Moser-Tardos def.]: a witness tree is a pair $\tau = (\tilde{t}, \tilde{\sigma})$ where $\tilde{t}$ is a unlabeled rooted tree and $\tilde{\sigma} : V_{\tilde{t}} \to X$ is a labeling such that: chidren labels are incompatible with the father label's and are distinct.

[Equivalent def.] Fix a total order in X. A witness tree is a pair $\tau = (t, \sigma)$ where t is a plane rooted tree (children are ordered) and $\sigma : V_\tau \to X$ is a *good labeling*, i.e. such that children labels are incompatible with the father label's and *respect the order*. [i.e. if v and w are siblings vertices in t and $v < w$ then $\sigma(v) < \sigma(w)$ (according to the order introduced in X)].

As the algorithm runs, resampling at each step some bad event from the family $\boldsymbol{A}$, *the Log of the algorithm* $C = \{C(1), C(2), \ldots\}$ with $C(i) \in X$ lists the events as they are resampled by the algorithm at each step, so that, for $i \in \mathbb{N}$, if $C(i) = x$ then the event $A_x \in \boldsymbol{A}$ resampled at step i of the algorithm.

C is a random variable determined by the random choices made by the algorithm at each step. If the algorithm stops then C is partial.

Moser Tardos associate to each step s of the algorithm, with log C, a witness tree $\tau_s = (t_s, \sigma_s)$ with root labeled $C(s)$ and the rest of labels in $\{C(j)\}_{j<s}$.

The tree τ_s is obtained by constructing a sequence $\tau_s^s, \tau_s^{s-1}, \ldots, \tau_s^1$ of witness trees [i.e. $\tau_s^i = (t_s^i, \sigma_s^i)$] and posing $\tau_s = \tau_s^1$.

The scheme goes as follows

A) τ_s^s is the tree formed only by a single vertex (i.e. the root) with label $C(s)$.

B) τ_s^{i-1} is obtained from $\tau_s^i = (t_s^i, \sigma_s^i)$ as follows: let $W_i = \{v \in t_s^i : \sigma_s^i(v) \nsim C(i-1)\}$.

• if $W_i = \emptyset$ then put $\tau_s^{i-1} = \tau_s^i$

• if $W_i \neq \emptyset$, then τ_s^{i-1} is obtained from τ_s^i by attaching a new vertex w to t_s^i with label $C(i-1)$ s.t. w is the child of a vertex $u \in W_i$ having the maximum distance from the root [if there is more than one of such vertices choose the one with maximum label (in the order of X]. *

Let $\mathcal{T}_x$ be the set os all possible distinct witness trees with root label x generated by the algorithm

*Of course, in order to obtain a good labeling of $\tilde{t}_s^{i-1}$, if the vertex u had already children in t_s^i (so that w becomes a new sibling of these children of u) attach the new vertex w with label $C(i-1)$ respecting the order of the children of u.

Moser and Tardos then prove that

-the probability $Prob(\tau)$ to see a witness tree $\tau = (t, \sigma)$ in the log C of the algorithm with vertex set V_t and labels $\{\sigma(v)\}_{v \in V_t}$ is at most

$$Prob(\tau) \leq \prod_{v \in V_t} Prob(A_{\sigma(v)}) \equiv \prod_{v \in V_\tau} p_{\sigma(v)} \tag{1.9}$$

Now, let N_x be the random variable that counts how many times the event A_x is resampled during the execution of the algorithm. I.e. N_x is, by definition, the number of occurrences of the event A_x in the log C of the algorithm, i.e. N_x is the number of *distinct* proper witness trees appearing in the log C that have their root labeled x.

Therefore one can bound the expectation of N_x simply by summing the bounds (1.9) on the probabilities $Prob(\tau)$ as τ varies in the set of the different witness trees with root labeled x. Thus the expected value $E(N_x)$ of N_x is bounded as

$$E(N_x) \leq \Phi_x(\boldsymbol{p}) \tag{1.10}$$

where

$$\Phi_x(\boldsymbol{p}) = \sum_{(t,\sigma)\in\mathcal{T}_x} \prod_{v\in V_t} p_{\sigma(v)} \tag{1.11}$$

Moser and Tardos's conclude their proof by showing, via a Galton-Watson branching process argument, that the quantity $\Phi_x(\boldsymbol{p})$ defined in (1.11) is bounded by μ_x if probabilities $\{p_x\}_{x\in X}$ are such that conditions (1.7) are verified. So the algorithm stops after expected $\sum_{x\in X} \mu_x$ steps.

Let us now give the following

Definition 1 *A proper witness tree $\tau = (t, \sigma)$ is called a Penrose tree if the following occurs:*

(t1) *if two vertices v and v' are such that $d(v) = d(v')$, then $\sigma(v) \sim \sigma(v')$;*

(t2) *if two vertices v and v' are such that $d(v') = d(v) - 1$ and $v^* \prec v'$ (i.e. v' is an uncle of v which is below the father v^* of v), then $\sigma(v) \sim \sigma(v')$*

We denote by $\mathcal{P}_x$ the set of all Penrose trees $\tau = (t, \sigma)$ with root label x.

Proposition 8 *Let $\tau = (t, \sigma)$ be a proper witness tree and let C be the (random) log produced by the algorithm. If τ occurs in the log C, then τ is a Penrose tree. In other words $\mathcal{T}_x \subset \mathcal{P}_x$.*

Proof. By absurd, that v and v' are two vertices of τ at the same distance from the root, i.e. $d(v) = d(v')$ and that the label of v is incompatible with the label of v'. Suppose, without loss in generality, that v' has been attached after v. But then, since the label of v' is incompatible with the label of v, we have that $d(v') \geq d(v) + 1$ contrary to the hypothesis that $d(v') = d(v)$. So if $d(v) = d(v')$ then necessarily $\sigma(v) \sim \sigma(v')$.

Suppose now that v and v' are vertices of τ such that v' is an uncle of v who is below the father v^* of v

in the drawing of τ. We need to show that $\sigma(v) \sim \sigma(v')$. By absurd suppose that $\sigma(v) \not\sim \sigma(v')$. We have to consider two cases. First we suppose that v has been added after v' to form $\tau(s) = \tau$. Since $\sigma(v) \not\sim \sigma(v')$ and v' is below v^*, then, according to the deterministic rule described above, v cannot be attached to v^*: it must be attached to v' or to another uncle below v', contrary to the hypothesis that v is attached to v^*. Secondly, suppose that v' has been added after v. But then $d(v') \geq d(v) + 1$, contrary to the hypothesis that v' is uncle of v (and hence $d(v') = d(v) - 1$). $\square$

Due to Proposition 8 we have that the expected number of times an event $A_x \in \boldsymbol{A}$ is resampled by the MT-algorithm is bounded by

$$E(N_x) \leq \tilde{\Psi}_{x_0}(-\boldsymbol{\rho}) \doteq \sum_{\substack{\tau=(t,\sigma)\\ \tau\in\mathcal{P}_x}} \prod_{v\in V_t} p_{\sigma(v)} \tag{1.12}$$

Penrose trees in the hard-core gas

Definition: Labeled rooted trees. A labeled rooted tree ϑ is a tree with vertex set $V_\vartheta = \mathrm{I}_n^0$ and root 0.

Let T_n^0 denotes the set of all labeled rooted trees with vertex set I_n^0 which are rooted in 0 and let $\mathbb{T}_n^0$ the set os all plane rooted trees with $n+1$ vertices.

There is a natural map $m : T_n^0 \to \mathbb{T}_n^0$ which associates to each labeled rooted tree $\vartheta \in T_n^0$ a unique plane rooted tree $m(\vartheta) \in \mathbb{T}_n^0$. This unique plane rooted tree $m(\vartheta)$ is obtained by fixing the order of the children in each vertex of ϑ according with the order of their labels in I_n^0.

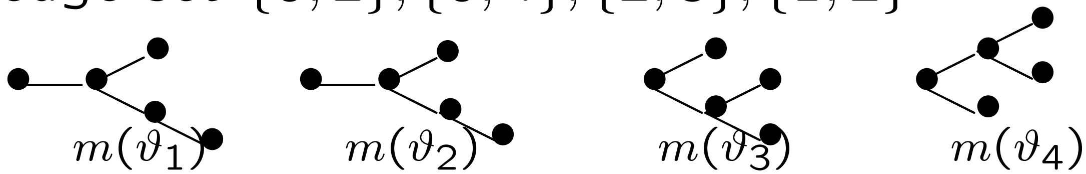
Example:

ϑ_1 with edge set $\{0,3\},\{1,3\},\{2,3\},\{2,4\}$,

ϑ_2 with edge set $\{0,2\},\{1,4\},\{2,3\},\{2,4\}$,

ϑ_3 with edge set $\{0,2\},\{0,3\},\{1,3\},\{3,4\}$

ϑ_4 with edge set $\{0,2\},\{0,4\},\{2,3\},\{1,2\}$



Remark. The map $\vartheta \mapsto m(\vartheta) = \tilde{t}$ is many-to-one and

$$\left|\{\vartheta \in T_n^0 : m(\vartheta) = \tilde{t}\}\right| = \frac{n!}{\prod_{v \in V_t} s_{v_i}!} \tag{1.13}$$

where s_v denotes the number of the children of the vertex v.

There is also a natural map $\theta : \mathbb{T}_n^0 \to T_n^0 : t \mapsto \vartheta_t$ (an injection) with the labeled tree ϑ_t such that:

- the root has label 0, the s_0 children of the root have labels $1, 2, \ldots, s_0$ from top to bottom, the higher root child vertex, i.e. that with label 1, has s_1 children with labels $s_0 + 1 \ldots s_0 + s_1$, the root child vertex with label i has s_i children with labels $s_0 + s_1 + \ldots s_{i-1} + 1, \ldots, s_0 + s_1 + \ldots s_{i-1} + s_i$, and so on. We call this labeling of t the *natural labeling* of a plane rooted tree t.

- This natural labeling for t induces a natural total order $\prec$ on the set of vertices V_t in a plane rooted tree $t \in \mathbb{T}_n^0$. I.e., given two (distinct) vertices u e v of t, we have $v \prec u$, if the natural label of v is less than the natural label of u. †

†In other words $v \prec u$ if either $d(v) < d(u)$, or $d(v) = d(u)$ but v is above u in the drawing of t [$d(v)$= distance between v and the root].

Remark. The total order introduced on the vertices of a plane rooted tree $t \in \mathbb{T}_n^0$ automatically induces a total order $\prec$ also on vertices of a labeled rooted tree $\tau \in T_n^0$.

Namely, given any two vertices u', v' in $\vartheta \in T_n^0$ we say that $u' < v'$ if the corresponding vertices u, v in $t = m(\vartheta) \in \mathbb{T}_n^0$ are such that $u \prec v$.

Obs: this total order of the vertices of a labeled rooted tree ϑ, which, we recall, are integers numbers, can be different from the standard order of the integers.

Notation. If v is a vertex in a rooted tree we denote by v^* the father of v.

The Penrose trees

Let's now go back to the graph $G = (X, E)$ in which the hard core lattice gas has been defined.

Definition 2 *The pair* $(\vartheta; (x_0, x_1, \ldots, x_n))$ *where* $\vartheta \in T_n^0$ *and* $(x_0, x_1, \ldots, x_n) \in X^{n+1}$ *is called a Penrose tree if the following holds.*

- *if* $\{i,j\} \in E_\vartheta$ *then* $\{i,j\} \in E_{g(x_0,x_1,\ldots,x_n)} \iff x_i \nsim x_j$

- *if two vertices* i *and* j *are such that* $d(i) = d(j)$, *then* $\{i,j\} \notin E_{g(x_0,x_1,\ldots,x_n)} \iff x_i \sim x_j$;

- *if two vertices* i *and* j *are such that* $d(j) = d(i) - 1$ *and* $i^* \prec j$, *then* $\{i,j\} \notin E_{g(x_0,x_1,\ldots,x_n)}$ *(i.e.* $x_i \sim x_j$*).*

We denote by $P(x_0, x_1, \ldots, x_n)$ the subset of T_n^0 constituted by those $\vartheta \in T_n^0$ such that the pair $(\vartheta; (x_0, x_1, \ldots, x_n))$ is Penrose.

Remark. Property (t0) says that (labels of) children always overlap (labels of) their parents, property (t1) says that siblings and/or cousins do not overlap. Finally property (t3) says that children are always compatible with their uncles which are below the father in the drawing of the plane tree $m(\vartheta)$. We want to emphasize that the map presented above is slightly different respect to the original map given by Penrose. The present definition has the advantage to be independent of the (integer) labels of the tree $\vartheta \in T_n^0$. It depends only on the underlying plane rooted tree $t = m(\vartheta)$

One can show the following

Proposition 9 (Penrose identity)

$$\Big|\sum_{\substack{g\in G_n^0\\ g\subset g(x_0,x_1,\dots,x_n)}}(-1)^{|E_g|}\Big| = \sum_{\vartheta\in T_n^0}\mathbb{1}_{\vartheta\in P(x_0,x_1,\dots,x_n)}(\vartheta) \tag{1.14}$$

where $\mathbb{1}_{\vartheta\in P(x_0,x_1,\dots,x_n)}$ *is the characteristic function of the set* $P(x_0,x_1,\dots,x_n)$ *in* T_n^0*, i.e.*

$$\mathbb{1}_{\vartheta\in P(x_0,x_1,\dots,x_n)}(\vartheta) = \begin{cases} 1 & \textit{if } \vartheta\in P(x_0,x_1,\dots,x_n) \\ \\ 0 & \textit{otherwise} \end{cases}$$

Using the Penrose identity we can rewrite the formal series (1.4) as

$$\Psi_{x_0}(-\boldsymbol{\rho}) = \rho_{x_0} \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{\vartheta \in T_n^0} \phi_{x_0}(\vartheta, \boldsymbol{\rho}) \tag{1.15}$$

where

$$\phi_{x_0}(\vartheta, \boldsymbol{\rho}) = \sum_{(x_1,\ldots,x_n)\in X^n} 1\!\!1_{\vartheta \in P(x_0,x_1,\ldots,x_n)} \, \rho_{x_1} \cdots \rho_{x_n} \tag{1.16}$$

Remark. The factor $\phi_{x_0}(\vartheta, \boldsymbol{\rho})$ only depends on the plane rooted tree associated to ϑ by the map m defined above. I.e., if $m(\vartheta) = t \in \mathbb{T}_n^0$

$$\phi_{x_0}(\vartheta, \boldsymbol{\rho}) = \phi_{x_0}(\vartheta_t, \boldsymbol{\rho}) \tag{1.17}$$

recall: ϑ_t is the natural labeled tree associated to t.

Therefore

$$\Psi_{x_0}(-\boldsymbol{\rho}) \ = \ \rho_{x_0}\sum_{n\geq 0}\frac{1}{n!}\sum_{t\in\mathbb{T}^0_n}\sum_{\substack{\vartheta\in T^0_n\\ m(\vartheta)=t}}\phi_{x_0}(\vartheta_t,\boldsymbol{\rho}) =$$

$$\Psi_{x_0}(-\boldsymbol{\rho}) \ = \ \rho_{x_0}\sum_{n\geq 0}\frac{1}{n!}\sum_{t\in\mathbb{T}^0_n}\phi_{x_0}(\vartheta_t,\boldsymbol{\rho})\sum_{\substack{\vartheta\in T^0_n\\ m(\vartheta)=t}} 1 =$$

i.e. we get

$$\Psi_{x_0}(-\boldsymbol{\rho}) =$$

$$\rho_{x_0}\sum_{n\geq 0}\sum_{t\in\mathbb{T}^0_n}[\prod_{v\in V_t}\frac{1}{s_v!}]\sum_{(x_1,\ldots,x_n)\in X^n}\mathbb{1}_{\vartheta_t\in P(x_0,x_1,\ldots,x_n)}\prod_{i=1}^{n}\rho_{x_i} \tag{1.18}$$

Remark. Note that Definition 1 coincides, *mutatis mutandis*, with Definition 2 given in section 2 in the following sense. If $\tau = (t, \sigma)$ is a Penrose tree according to definition 1, then t, being a plane rooted tree, defines uniquely the labeled rooted tree $\vartheta_t \in T_n^0$ previously seen. Moreover the functi on σ defines uniquely a $n+1$-tuple $(x_0, x_1, \ldots, x_n)$ such that $\sigma(i) = x_i$ for each $i \in I_n^0$ (we are identifying vertices of V_t with numbers in I_n^0 through the bijection $t \mapsto \vartheta_t$). Then $\vartheta_t \in P(x_0, x_1, \ldots, x_n)$ according to definition 2. I.e.

$$\tau = (t, \sigma) \in \mathcal{P}_x \quad \Longleftrightarrow \quad \vartheta_t \in P(x_0, x_1, \ldots, x_n)$$

Therefore recalling formulas (1.18) we get

$$\Psi_{x_0}(-\boldsymbol{\rho}) =$$

$$\rho_{x_0} \sum_{n\geq 0} \sum_{t\in\mathbb{T}_n^0} [\prod_{v\in V_t} \frac{1}{s_v!}] \sum_{(x_1,\ldots,x_n)\in X^n} \mathbb{1}_{\vartheta_t \in P(x_0,x_1,\ldots,x_n)} \prod_{i=1}^{n} \rho_{x_i} =$$

$$= \rho_{x_0} \sum_{\substack{\tau=(t,\sigma) \\ \tau\in\mathcal{P}_x}} \prod_{v\in V_t} \rho_{\sigma(v)}$$

and due to Proposition 8 we have that the expected number of times an event $A_x \in \boldsymbol{A}$ is resampled by the MT-algorithm is bounded by

$$E(N_x) \leq \tilde{\Psi}_{x_0}(-\boldsymbol{\rho}) \doteq \sum_{\substack{\tau=(t,\sigma) \\ \tau\in\mathcal{P}_x}} \prod_{v\in V_t} p_{\sigma(v)} \qquad (1.19)$$

which concludes the proof of Theorem 6.

Proof of Proposition 8.

Fix $(x_0, x_1, \dots, x_n) \in X^{n+1}$. Then is uniquely defined the $g(x_0, x_1, \dots, x_n)$ with vertex set I_n^0 and edge set $E_{g(x_0,x_1,\dots,x_n)} = \{\{i,j\} \subset \mathrm{I}_0^n : \ x_i \nsim x_j\}$. Without loss in generality we may assume that $g(x_0, x_1, \dots, x_n)$ is connected (otherwise $\phi^T(x_0, x_1, \dots, x_n) = 0$ and (1.14) is trivial). We denote by G_n^0 the set of all connected graphs with vertex set I_n^0 and we put

$$G_{g(x_0,x_1,\dots,x_n)} = \{g \in G_n^0 : g \subset g(x_0, x_1, \dots, x_n)\}$$

and

$$T_{g(x_0,x_1,\dots,x_n)} = \{\vartheta \in T_n^0 : \vartheta \subset g(x_0, x_1, \dots, x_n)\}$$

Let us define the map $q : G_{g(x_0,x_1,\dots,x_n)} \to T_{g(x_0,x_1,\dots,x_n)}$ that associate to $g \in G_{g(x_0,x_1,\dots,x_n)}$ a unique labeled rooted tree $q(g) \in T_{g(\gamma_0,\gamma_1,\ldots,\gamma_n)}$ as follows. We recall that the vertices of $g \in G_{g(x_0,x_1,\ldots,x_n)}$ are labeled with labels in $\{0, 1, 2, \ldots, n\}$ and we are denoting by E_g the edge set of g. We also consider the graph g as always rooted in 0, so for any j vertex of g, we will denote by $d_g(j)$ its distance from the root 0 in g.

1) We first delete all edges $\{i, j\}$ in E_g with $d_g(i) = d_g(j)$. After this operation we are left with a connected graph g' such that $d_{g'}(i) = d_g(i)$ for all vertices $i = 0, 1, \ldots, n$. Moreover each edge $\{i, j\}$ of g' is such that $|d_{g'}(i) - d_{g'}(j)| = 1$.

2) Let $i_1, \ldots, i_{s_0}$ be the vertices at distance 1 from the root 0 in g' ordered in such way that $i_1 < i_2 < \cdots < i_{s_0}$

(note that we identify vertices with their labels, so that $\{i_1, \ldots, i_{s_0}\}$ is a subset $\{0, 1, 2, \ldots, n\}$). Now take the smaller of these vertices, say i_1, and let $j_1^{i_1}, \ldots, j_{s_{i_1}}^{i_1}$ be the vertices connected to i_1 by edges of $E_{g'}$ (these vertices are at distance 2 from the root 0 and again are ordered according their labels) and delete all edges of g' connecting vertices $j_1^{i_1}, \ldots, j_{s_{i_1}}^{i_1}$ to vertices in the set $\{i_2, \ldots, i_{s_0}\}$. The graph so obtained g'_1 is such that any of the vertices $j_1^{i_1}, \ldots, j_{s_{i_1}}^{i_1}$ is connected only to i_1 and vertices at distance greater than 2. Then take the vertex i_2 (the smaller after i_1) and let $j_1^{i_2}, \ldots, j_{s_{i_2}}^{i_2}$ be the vertices connected to i_2 at distance 2 from the root 0 in g'_1 and delete all edges of g'_1 connecting vertices $j_1^{i_2}, \ldots, j_{s_{i_2}}^{i_2}$ to vertices in the set $\{i_3, \ldots, i_{s_0}\}$. The graph so obtained g'_2 is such that any of the vertices

$j_1^{i_1}, \ldots, j_{s_{i_1}}^{i_1}$ is connected only to i_2 and vertices at distance greater than 2. After s_0 steps we are left with a graph g'_{s_0} with no loops among vertices at distance $d \leq 2$ from the root. Continue now this procedure until all vertices of g are exhausted, always respecting the order of the labels. Namely, take $j_1^{i_1}$ (i.e. the one with the smaller label among $j_1^{i_1}, \ldots, j_{s_{i_1}}^{i_1}$) and consider the vertices at distance 3 emanating from $j_1^{i_1}$ and delete all edges linking these vertices to some vertex in the set $\{j_2^{i_1}, \ldots, j_{s_{i_1}}^{i_1}, j_1^{i_2}, \ldots, j_{s_{i_2}}^{i_2}, \ldots, j_1^{i_{s_0}}, \ldots, j_{s_{i_{s_0}}}^{i_2}\}$ and continue this procedure until all vertices are exhausted. The resulting graph $g'' \doteq q(g)$ is by construction a spanning connected subgraph of $g(x_0, x_1, \ldots, x_n)$, i.e. $q(g) \in G_{g(x_0,x_1,\ldots,x_n)}$, and which has no cycles, i.e. $q(g) \in T_{g(x_0,x_1,\ldots,x_n)}$. Observe that the map q is a surjection from $G_{g(x_0,x_1,\ldots,x_n)}$ to $T_{g(x_0,x_1,\ldots,x_n)}$.

Conversely, Let p be the map that to each tree $\vartheta \in T_{G(x_0,x_1,\ldots,x_n)}$ associates the graph $p(\vartheta) \in G_{g(x_0,x_1,\ldots,x_n)}$ formed by adding to ϑ all edges $\{i,j\} \in E_{g(x_0,x_1,\ldots,x_n)} \setminus E_\vartheta$ such that either $d_\vartheta(i) = d_\vartheta(j)$, or $d_\vartheta(j) = d_\vartheta(i) - 1$ and $i^* \prec j$.

Observe now that the set $G_{g(x_0,x_1,\ldots,x_n)}$ is partially ordered by edge inclusion, namely, $g, g \in G_{g(x_0,x_1,\ldots,x_n)}$ and $E_g \subset E_{g'}$, then $g < g'$. Moreover if $g, g' \in G_{g(x_0,x_1,\ldots,x_n)}$ and $g < g'$ we denote by $[g, g']$ the subset of $G_{g(x_0,x_1,\ldots,x_n)}$ formed by those $\widehat{g}$ such that $g < \widehat{g} < g'$. With these definitions we have that if $\vartheta \in T_{g(x_0,x_1,\ldots,x_n)}$ and $g \in [\vartheta, p(\vartheta)]$, then, by construction of the map m, we have that $m(g) = \vartheta$, i.e., among those graphs $g \in G_{g(x_0,x_1,\ldots,x_n)}$ such that $q(g) = \vartheta$ ϑ is the minimal graph and $p(\vartheta)$ is the maximal graph, respect to the partial order relation $<$ in $G_{g(x_0,x_1,\ldots,x_n)}$. So $G_{g(x_0,x_1,\ldots,x_n)}$ is partitioned in the disjoint union of the sets $[\vartheta, p(\vartheta)]$ with $\vartheta \in T_{g(x_0,x_1,\ldots,x_n)}$. This shows that the map p provides a so-called *partition scheme* of the family of graphs $G_{g(x_0,x_1,\ldots,x_n)}$. Observe finally, recalling Definition 2, that if $\vartheta \in T_{g(x_0,x_1,\ldots,x_n)}$, then $p(\vartheta) = \vartheta \iff \vartheta \in P(x_0, x_1, \ldots, x_n)$.

With these definitions we have

$$\sum_{g \in G_{g(x_0,x_1,\ldots,x_n)}} (-1)^{|E_g|} =$$

$$= \sum_{\vartheta \in T_{g(x_0,x_1,\ldots,x_n)}} (-1)^{|E_\vartheta|} \sum_{\substack{g \in G_{g(x_0,x_1,\ldots,x_n)} \\ q(g)=\vartheta}} (-1)^{|E_g|-|E_\vartheta|} =$$

$$= (-1)^n \sum_{\vartheta \in T_{g(x_0,x_1,\ldots,x_n)}} [1 + (-1)]^{|E_{p(\vartheta)}| - |E_\vartheta|} =$$

$$= (-1)^n \sum_{\substack{\vartheta \in T_{g(x_0,x_1,\ldots,x_n)} \\ p(\vartheta)=\vartheta}} 1 =$$

$$= (-1)^n \sum_{\vartheta \in T_n^0} \mathbb{1}_{\vartheta \in P(x_0,x_1,\ldots,x_n)}$$

and the proposition is proved. □