Counting and Sampling **Fall 2017**

Lecture 12: Correlation Decay: Matching Polynomial

Lecturer: Shayan Oveis Gharan *Nov 3rd*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

In this lecture we give a FPTAS to estimate the value of the matching polynomial of a given graph at a given point. For a graph G, the (positive) matching polynomial is defined as follows:

$$\mu_G^+(x) = \sum_k m_k(G)x^k \tag{12.1}$$

where $m_k(G)$ is the number of k-matchings of G. Note that the matching polynomial is conventionally defined as

$$\mu_G(x) := \sum_k (-1)^k m_k(G)x^{2n-k} = x^{2n} - m_1(G)x^{2n-2} + \dots \tag{12.2}$$

But in counting applications it is more convenient to work with (12.1). For example, for $x = 1$, (12.1) counts the number of *all* matchings of G. Note that the polynomials are closely related. As we will see in future $\mu_G(x)$ is real rooted; so if $\pm r_1, \dots, \pm r_n$ are the roots of $\mu_G(x)$, then $-1/r_1^2, \dots, -1/r_n^2$ are the roots of $\mu_G^+(x)$. We will say more about the roots in future.

For simplicity of notation in this lecture we write $\mu_G(x)$ to denote $\mu_G^+(x)$. Here, we prove the following theorem.

Theorem 12.1 ([BGK+07])**.** *There is a an algorithm that for any given $\epsilon > 0$ and a graph G with degree at most Δ and any bounded t estimates $\mu_G(t)$ within $1 + \epsilon$ multiplicative factor in time polynomial in $n, 1/\epsilon$. The exponent depends linearly on t, Δ.*

Fix a number $t > 0$, We can think of a probability distribution over all matchings of G where for every matching M, $\mathbb{P}[M] \propto t^{|M|}$, i.e.,

$$\mathbb{P}[M] = \frac{t^{|M|}}{\mu_G(t)}.$$

We abuse notation and also call this probability distribution $\mu_G(t)$. Let us first discuss a few observations about this probability distribution. Obviously, $\mu_G(t)$ is the partition function of this distribution. For any vertex $v \in V$,

$$\mu_{G,v}(t) := \frac{\mu_{G-v}(t)}{\mu_G(t)} = \mathbb{P}_{M \sim \mu_G(t)}[v \text{ is not saturated in } M].$$

We will design an algorithm to estimate the above quantity. Note that if we can approximate the above quantity up to a $(1 + \epsilon/n)$ error, then we can estimate the partition function $\mu_G(t)$ up to $1 + \epsilon$ error

Lemma 12.2. *Given an oracle and $t > 0$ such that for any graph G and a vertex v estimates $\mu_{G,v}(t)$ up to a $1 + \epsilon/2n$ multiplicative error for $n = |V|$, we can estimate $\mu_G(t)$ up to a $1 + \epsilon$ multiplicative error.*

Proof. The lemma simply follows from the telescopic product:

$$\begin{aligned}\mu_G(t) &= \left(\frac{\mu_{G-v_1}(t)}{\mu_G(t)} \cdot \frac{\mu_{G-v_1-v_2}(t)}{\mu_{G-v_1}(t)} \cdot \dots \cdot \frac{1}{\mu_{G-v_1-\dots-v_{n-1}}(t)}\right)^{-1} \\ &= \left(\mu_{G,v_1}(t) \cdot \mu_{G-v_1,v_2}(t) \dots \mu_{G-v_1-\dots-v_{n-1},v_n}(t)\right)^{-1}.\end{aligned}$$

Using the oracle we can estimate each of the terms in the RHS up to a $(1+\epsilon/2n)$ error. So, we get a $(1+\epsilon/2n)^n$ approximation of $1/\mu_G(t)$ whose inverse a $1+\epsilon$ approximation of $\mu_G(t)$. □

Another related quantity is

$$t\frac{d}{dt}\ln\mu_G(t) = t\frac{\mu'_G(t)}{\mu_G(t)} = \sum_k k\cdot m_k(G)\frac{t^k}{\mu_G(t)} = \mathbb{E}_{M\sim\mu_G(t)}\left[|M|\right].$$

12.1 Correlation Decay

The main idea in correlation decay counting algorithms is to build a recursive procedure for computing the marginal probability that any given vertex is in solution; for example a given vertex is saturated in the random matching sampled from $\mu_G(t)$. The recursion works by examining sub-instances with "boundary conditions" which require certain vertices to be in, or out, of the independent set. The recursion structure is called a *computation tree*. Nodes of the tree correspond to intermediate instances, and boundary conditions are different in different branches. The computation tree allows one to compute the marginal probability exactly but the time needed to do so may be exponentially large since, in general, the tree is exponentially large. Typically, an approximate marginal probability is obtained by truncating the computation tree to logarithmic depth so that the (approximation) algorithm runs in polynomial time. If the correlation between boundary conditions at the leaves of the (truncated) computation tree and the marginal probability at the root decays exponentially with respect to the depth, then the error incurred from the truncation is small and the algorithm succeeds in obtaining a close approximation.

First, we prove the weak spatial mixing property of the underlying distribution of matchings for small enough t and then we see how to use this to count matchings.

We start by proving natural properties of matching polynomial

Lemma 12.3. *The following properties of the matching polynomial are immediate from the definition.*

i) For any disjoint graphs G, H,

$$\mu_{G\cup H}(x) = \mu_G(x)\mu_H(x).$$

ii) For any graph $G=(V,E)$ and any vertex $u\in V$,

$$\mu_G(x) = \mu_{G-u}(x) + x\sum_{v\sim u}\mu_{G-u-v}(x).$$

Proof. Property (i) follows from the fact that the union of any matching $M\in G$ and $M'\in H$ is a matching $M\cup M'\in G\cup H$.

Property (ii) follows from the fact that for any matching $M\in G$ either vertex u is not saturated in which case M is counted in the polynomial $\mu_{G-u}(x)$. Otherwise, u is matched to one of its neighbors, say v. In the latter case $M-(u,v)$ is a matching in $G-u-v$. Obviously, this argument counts every matching of G exactly once. □

Building on (ii), for a vertex $u\in V$ we can write the marginal probability of u in the distribution $\mu_G(t)$ as follows:

$$\frac{\mu_{G-u}(t)}{\mu_G(t)} = \frac{1}{1+t\sum_{v\sim u}\frac{\mu_{G-u-v}(t)}{\mu_{G-u}(t)}} = \frac{1}{1+t\sum_{v\sim u}\mu_{G-u,v}(t)}. \tag{12.3}$$

Having this, we can think of a computation tree to approximate the marginal probability of u. In particular, the value of every node of the computation tree is the marginal probability of a node v in a graph H, $\mu_{H,v}(t)$ where H is a subgraph of G obtained by deleting a number of nodes of G. The children of such a node correspond to the marginal probabilities $\mu_{H-v,w}(t)$ for every neighbor w of v in H.

Of course, the depth of the computation tree is n, and if we go all the way down the tree we can compute the marginal probabilities exactly. But here the main observation is that the marginal probability at the root is almost independent of the nodes far enough down tree.

The following lemma is our key tool. Suppose we have a very bad estimate of the marginal probability of the nodes at depth k of the computation tree. For simplicity, we can assume all probabilities are 1, call this vector $\mathbf{x}$, and let $\mathbf{y}$ be the true marginal probabilities of nodes at that depth. Given $\mathbf{x}$ (or $\mathbf{y}$) can use (12.3) to calculate marginal probabilities of all nodes all the way to root. We show that no matter how far $\mathbf{x}$ is from $\mathbf{y}$ as we get closer and closer to the root $\|\mathbf{x}-\mathbf{y}\|_\infty$ goes down geometrically.

In the following lemma we show that

Lemma 12.4. *For any graph H and any vertex $v \in H$ and any two vectors $\mathbf{x}, \mathbf{y} \le 1$*

$$\left| \ln \frac{1}{1+t\sum_{w\sim v} x_{H-v,w}} - \ln \frac{1}{1+t\sum_{w\sim v} y_{H-v,w}} \right| \le \frac{t\Delta}{t\Delta+1} \|\ln \mathbf{x} - \ln \mathbf{y}\|_\infty$$

where d is the degree of v.

Proof. Also, let $a = \ln x$ and $b = \ln y$. We can rewrite the claim as follows:

$$\left| \ln(1+t\sum_{w\sim v} e^{a_{H-v,w}}) - \ln(1+t\sum_{w\sim v} e^{b_{H-v,w}}) \right| \le \frac{t\Delta}{1+t\Delta} \|\mathbf{a}-\mathbf{b}\|_\infty.$$

Note that expectations in the LHS are with respect to a uniform distribution on neighbors of v. For a vector $\mathbf{z}$ let $f(\mathbf{z}) = \ln(1+t\sum_{w\sim v} e^{z_{H-v,w}})$. So, it is enough to upper bound $|f(\mathbf{a}) - f(\mathbf{b})|$. By the mean value theorem there exists a point $\mathbf{z} = \alpha\mathbf{a} + (1-\alpha)\mathbf{b}$ on the line connecting $\mathbf{a}$ to $\mathbf{b}$ such that

$$|f(\mathbf{a}) - f(\mathbf{b})| = |\langle \nabla f(\mathbf{z}), (\mathbf{a}-\mathbf{b})\rangle| \le \|\nabla f(z)\|_1 \cdot \|\mathbf{a}-\mathbf{b}\|_\infty$$

where the inequality follows by the Holder inequality. So, to prove the lemma it is enough to show that $\|\nabla f(z)\|_1 \le td/(1+td)$. We have

$$\begin{aligned} \|\nabla f(z)\|_1 &= \sum_{w\sim v} \frac{te^{z_{H-v,w}}}{1+t\sum_{w\sim v} e^{z_{H-v,w}}} \\ &= 1 - \frac{1}{1+t\sum_{w\sim v} e^{z_{H-v,w}}} \\ &\le 1 - \frac{1}{1+t\cdot\deg(v)} \le \frac{t\Delta}{1+t\Delta} \end{aligned}$$

where the first inequality follows by the fact that for any w, $e^{a_{H-v,w}} = x_{H-v,w}, e^{b_{H-v,w}} = y_{H-v,w} \le 1$. So, $a_{H-v,w}, b_{H-v,w} \le 0$ which implies $z_{H-v,w} \le 0$ and $e^{z_{H-v,w}} \le 1$. The last inequality simply follows from $\deg(v) \le \Delta$. □

To finish the proof of Theorem 12.1 we need to take care of the following concerns:

- At the beginning some coordinates of the correct marginal vector, $\mathbf{y}$, may be zero. This implies that $\ln \mathbf{y}$ can be arbitrarily large. So, the above lemma seems useless in such a case.

As we show in the following lemma, after one step all marginal probabilities will be in the interval $[1/1+t\Delta, 1]$. This implies that for an arbitrary vector $\mathbf{x}$ and the true marginal vector $\mathbf{y}$, in one step $\|\ln \mathbf{x} - \ln \mathbf{y}\| \leq \ln(1+t\Delta)$.

- Also, by the following lemma even if we start from a wrong marginal vector, the estimate is always at most 1, so the assumptions of Lemma 12.4 are always satisfied.
- We have a good estimate of the marginals at the root when $(1-\epsilon/2n)y_{G,v} \leq x_{G,v}(1+\epsilon/2n)y_{G,v}$. This means that
$$|\ln x_{G,v} - \ln y_{G,v}| \leq |\ln(1-\epsilon/2n)| \leq \epsilon/2n.$$
So, it is enough to start the estimation process at depth $O(t\Delta \log(n/\epsilon))$.

The following lemma gives a few facts about the above recursion process.

Lemma 12.5. *For any graph H, vertex v of degree d, and a vector $0 \leq \mathbf{x} \leq 1$ we have*
$$\frac{1}{1+td} \leq \frac{1}{1+t\sum_{w\sim v} x_{H-v,w}} \leq 1$$

Proof. Let
$$f(\mathbf{x}) = \frac{1}{1+t\sum_{w\sim v} x_{H-v,w}}.$$
Observe that $\frac{1}{1+tx}$ is a decreasing function of x. Therefore, the maximum of $f(\mathbf{x})$ is obtained at $\mathbf{0}$, in which case
$$f(\mathbf{x}) \leq f(\mathbf{0}) = 1.$$
On the other hand, the minimum of $f(\mathbf{x})$ is obtained at $\mathbf{1}$ in which case
$$f(\mathbf{x}) \geq f(\mathbf{1}) = \frac{1}{1+td}.$$
□

This completes the proof of Theorem 12.1.

References

[BGK+07] Mohsen Bayati, David Gamarnik, Dimitriy A. Katz, Chandra Nair, and Prasad Tetali. Simple deterministic approximation algorithms for counting matchings. In *STOC*, pages 122–127, 2007. 12-1