Approximate Counting and Mixing Time of Markov chains	Fall 2024

Lecture 1: Max Entropy Programs and Entropy Contraction

Lecturer: Shayan Oveis Gharan 01/04/22

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

1.1 Newton Polytope

Given a polynomial

$$p(z_1, \ldots, z_n) = \sum_{\kappa \in \mathbb{Z}_{\geq 0}} c_p(\kappa) z^\kappa,$$

where $c_p(\kappa)$ is the coefficient of z^κ in p, the *Newton polytope* of p is the convex hull of all integer vectors κ with non-zero coefficient,

$$\text{Newt}(p) := \text{conv}\{\kappa \in \mathbb{Z}_{\geq 0} : c_p(\kappa) \neq 0\}$$

For example, if p is the generating polynomial of all spanning trees of a graph G, $\sum_T z^T$, then Newt(p) is the spanning tree polytope of G, the convex hull of the indicator vectors of all spanning trees of G.

Lemma 1.1. *For any polynomial $p \in \mathbb{R}_{\geq 0}[z_1, \ldots, z_n]$, and any $\alpha \in \mathbb{R}^n_{\geq 0}$, we have $\inf_{z>0} \frac{p(z)}{z^\alpha} > 0$ iff $\alpha \in$ Newt(p).*

Proof. $\Leftarrow$: First, assume that $\alpha \in$ Newt(p). Then, there is a convex combination of the vertices of this polytope that is equal to α,

$$\alpha = \sum_{\kappa : c_p(\kappa) \neq 0} \lambda_\kappa \kappa$$

where $\sum_\kappa \lambda_\kappa = 1$ and each $\lambda_\kappa \geq 0$. Then, for any $z > 0$ we can write,

$$p(z) = \sum_{\kappa \in \mathbb{Z}^n_{\geq 0}} \lambda_\kappa \frac{c_p(\kappa) z^\kappa}{\lambda_\kappa} \geq \prod_{\kappa \in \mathbb{Z}^n_{\geq 0}} \left(\frac{c_p(\kappa) z^\kappa}{\lambda_\kappa}\right)^{\lambda_\kappa} = z^\alpha \prod_{\kappa \in \mathbb{Z}^n_{\geq 0}} \left(\frac{c_p(\kappa)}{\lambda_\kappa}\right)^{\lambda_\kappa},$$

where the inequality follows by the weighted AM-GM inequality and that $c_p(\kappa) \geq 0$ and $z > 0$. Therefore, $\inf_{z>0} \frac{p(z)}{z^\alpha} \geq \prod_{\kappa \in \mathbb{Z}^n} \left(\frac{c_p(\kappa)}{\lambda_\kappa}\right)^{\lambda_\kappa} > 0$ as desired.

$\Rightarrow$: Conversely, suppose $\alpha \notin$ Newt(p). Then, there exists a separating hyperplane, i.e., there exists $c \in \mathbb{R}^n$ such that $\langle c, \alpha \rangle > b$ and $\langle c, x \rangle \leq b$ for any $x \in$ Newt(p) for some $b \in \mathbb{R}$. Suppose $\langle c, \alpha \rangle \geq b + \epsilon$ for some $\epsilon > 0$. Now, let $z^* = \exp(tc)$ where $t > 0$ is a sufficiently large number. Then,

$$\begin{aligned}
\inf_{z>0} \frac{p(z)}{z^\alpha} &\leq \frac{p(z^*)}{{z^*}^\alpha} \\
&= \frac{\sum_{\kappa \in \mathbb{Z}^n_{\geq 0}} c_p(\kappa) e^{\langle \log z^*, \kappa \rangle}}{e^{\langle \log z^*, \alpha \rangle}} \\
&= \frac{\sum_{\kappa \in \mathbb{Z}^n_{\geq 0}} c_p(\kappa) \exp(t\langle c, \kappa \rangle)}{\exp(t\langle c, \alpha \rangle)} \leq \frac{\sum_{\kappa \in \mathbb{Z}^n_{\geq 0}} c_p(\kappa) \exp(tb)}{\exp(t(b+\epsilon)}
\end{aligned}$$

Letting $t \to \infty$ the RHS converges to 0. □

1.2 Capacity Functions and Dual of Max Entropy Programs

Next, we prove the following theorem:

Theorem 1.2. *Let $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be a probability distribution. Let $\alpha \in Newt(g_\mu)$. Then, there exists an external field $(\lambda_1, \ldots, \lambda_n)$ such that for any $1 \leq i \leq n$,*

$$\mathbb{P}_{\lambda * \mu}[i] = \alpha_i,$$

i.e., the marginal probability of i under the distribution $\mu * \lambda$ is α_i.

The above theorem conceptually has a very important message. If μ is a family of probability distributions which are closed under taking external fields such as strongly Rayleigh, sector stable or log-concave distributions, then, given any point α in the Newton polytope of g_μ, there is *another distribution from the same family* μ' such that the marginals of μ' is equal to α.

Remark 1.3. *We remark that if α is in the interior of the Newton polytope we can attain α exactly, otherwise, we can only satisfy α as a marginal approximately, i.e., we can find a sequence of external field vectors $\lambda^1, \lambda^2, \ldots$ such that the marginal vectors of the distributions $\mu * \lambda^1, \mu * \lambda^2, \ldots$ converge to α.*

Recall that many of the probabilistic operations on μ can be translated to operations on the generating polynomial g_μ. To prove the theorem, it is natural to write down the marginal vector of a distribution μ: For any $1 \leq i \leq n$ we can write

$$\mathbb{P}_{S \sim \mu}[i \in S] = \partial_{z_i} g_\mu(z) \mid_{z=1}.$$

Sometimes, it is cleaner to assume g_μ is not normalized to $g_\mu(\mathbf{1}) = 1$. In such a case, we can write

$$\mathbb{P}_{S \sim \mu}[i \in S] = \frac{\partial_{z_i} g_\mu(z)}{g_\mu(z)}\Big|_{z=\mathbf{1}} = \partial_{z_i} \log g_\mu(z) \mid_{z=\mathbf{1}}. \tag{1.1}$$

We write the following convex program (which turns out to be the dual of the maximum entropy convex program):

$$\inf_y \log \frac{g_\mu(e^{y_1}, \ldots, e^{y_n})}{e^{\langle y, \alpha \rangle}}. \tag{Max-Entropy CP}$$

In particular, the above program can be obtained by a change of variables $z_i \leftrightarrow e^{y_i}$ and taking log of the objective value.

Since the above convex program has no constraints, the optimum solution is attained unless the optimum value is $-\infty$. In Lemma 1.1 we argued that the above infimum is $-\infty$ iff $\alpha \notin \text{Newt}(p)$. So, since $\alpha \in \text{Newt}(p)$, the infimum is bounded and we assume y^* is (an) optimum solution.

Since y^* is an optimal solution, the Gradient of the convex function must be zero at y^*; so for each $1 \leq i \leq n$ we can write

$$0 = \partial_{y_i} \left(\log g_\mu(e^{y_1}, \ldots, e^{y_n}) - \langle y, \alpha \rangle \right) \mid_{y=y^*}$$

Therefore,

$$\frac{\partial_{y_i} g_\mu(e^{y_1}, \ldots, e^{y_n})}{g_\mu(e^{y_1}, \ldots, e^{y_n})}\Big|_{y=y^*} = \alpha_i \tag{1.2}$$

Letting $\lambda = e^{y^*}$, i.e., $\lambda_i = e^{y_i^*}$ for all i, observe that

$$\begin{aligned} g_\mu(e^{y_1}, \ldots, e^{y_n})|_{y=y^*} &= g_\mu(\lambda_1 z_1, \ldots, \lambda_n z_n)_{z=\mathbf{1}}, \\ \partial_{y_i} g_\mu(e^{y_1}, \ldots, e^{y_n})|_{y=y^*} &= \partial_{z_i} g_\mu(\lambda_1 z_1, \ldots, \lambda_n z_n)|_{z=\mathbf{1}}. \end{aligned}$$

Therefore, by (1.1) for any $1 \leq i \leq n$,

$$\mathbb{P}_{S \sim \lambda * \mu}[i] = \partial_{z_i} \log g_{\lambda * \mu}(z)|_{z=\mathbf{1}} = \frac{\partial_{z_i} g_\mu(\lambda_1 z_1, \ldots, \lambda_n z_n)}{g(\lambda_1, \ldots, \lambda_n)}\Big|_{z=\mathbf{1}} = \alpha_i,$$

as desired. The last identity follows by (1.2)

(Max-Entropy CP) is called the maximum entropy convex program. This can be seen as a generalization of the convex program proposed by Gurvits that we discussed in Lecture 3. To computationally solve (Max-Entropy CP) we need to be able to evaluate the generating polynomial of μ and evaluate its partial derivatives. If μ is a strongly Rayleigh distribution, we can approximately evaluate g_μ. To be precise, one also needs to study the bit precision of the optimum solution y^*. It is a-priori unclear if the optimal solution y^* can be represented (or approximated) by polynomially (in n) many bits. This questions is well studied [**AGMOS17**; **SV14**; **SV19**] and it is not in the scope of this course.

1.3 Max entropy programs

Let $p \in \mathbb{R}_{\geq 0}[z_1, \ldots, z_n]$ and let $\alpha = \text{Newt}(p)$. Consider the following convex program:

$$\begin{aligned}
\max \quad & \sum_{\kappa \in \text{Newt}(p)} q_\kappa \log \frac{c_p(\kappa)}{q_\kappa} \\
\text{s.t.,} \quad & \sum_{\kappa \in \text{Newt}(p)} q_\kappa \kappa_i = \alpha_i \qquad \forall 1 \leq i \leq n, \\
& \sum_\kappa q_\kappa = 1 \\
& q_\kappa \geq 0 \qquad\qquad\qquad \forall \kappa.
\end{aligned} \tag{Max-Entropy Dual}$$

Claim 1.4. *The above program is the dual to* (Max-Entropy CP)*.*

We think of q as a distribution over integer points in $\text{Newt}(p)$. To write the dual of this program, we first need to write the Lagrangian:

$$\max_{q>0} \inf_{y \in \mathbb{R}^n} L(q, \gamma) = \max_{q>0} \inf_{y} \sum_{\kappa \in \text{Newt}(p)} q_\kappa \log \frac{c_p(\kappa)}{q_\kappa} - \sum_{i=1}^n y_i \left(\alpha_i - \sum_{\kappa \in \text{Newt}(p)} q_\kappa \kappa_i \right) - s \left(1 - \sum_{\kappa \in \text{Newt}(p)} q_\kappa \right)$$

By strong duality we can substitute the max and inf, so

$$\max_{q>0} \inf_{y \in \mathbb{R}^n, s} L(q, \gamma, s) = \inf_{y \in \mathbb{R}^n, s} \max_{q>0} L(q, y, s) \tag{1.3}$$

At optimality the gradient of the Lagrangian is zero, so for any κ,

$$\partial_{q_\kappa} L(q, y, s) = 0 \Leftrightarrow \log \frac{c_p(\kappa)}{q_\kappa} - 1 = -\sum_{i=1}^n y_i \kappa_i = -\langle y, \kappa \rangle - s.$$

Therefore, at optimality

$$\frac{c_p(\kappa)}{q_\kappa} = e^{1 - \langle y, \kappa \rangle - s}.$$

Plugging this into (1.3), we can write the dual as follows:

$$\inf_{y,s} \sum_{\kappa\in\text{Newt}(p)} q_\kappa(1-\langle y,\kappa\rangle - s) - \langle y,\alpha\rangle + \sum_{i=1}^n y_i \sum_{\kappa\in\text{Newt}(p)} q_\kappa \kappa_i - s + s\sum_{\kappa\in\text{Newt}(p)} q_\kappa \tag{1.4}$$

$$= \inf_{y,s} \sum_{\kappa\in\text{Newt}(p)} q_\kappa - \langle y,\alpha\rangle - s \tag{1.5}$$

$$= \inf_{y,s} \sum_{\kappa\in\text{Newt}(p)} c_p(\kappa) e^{s+\langle y,\kappa\rangle-1} - \langle y,\alpha\rangle - s \tag{1.6}$$

Optimizing the RHS over s we get

$$1 = \sum_{\kappa\in\text{Newt}(p)} c_p(\kappa)e^{s+\langle y,\kappa\rangle-1} \Leftrightarrow s = -\log \sum_{\kappa\in\text{Newt}(p)} c_p(\kappa) e^{\langle y,\kappa\rangle -1}$$

Plugging in the value of s, we can rewrite the dual as follows:

$$\inf_y 1 - \langle y,\alpha\rangle + \log \sum_{\kappa\in\text{Newt}(p)} c_p(\kappa)e^{\langle y,\kappa\rangle-1} = \inf_y \log \frac{p(e^{y_1},\dots,e^{y_n})}{y^\alpha}$$

as desired.

1.4 Entropy Contration

Lemma 1.5. *Suppose g_μ is log-concave. Then,*

$$g_\mu(z_1, ; z_n)^{1/k} \le \sum_{i=1}^n z_i \pi^1(i).$$

Note that in this proof we are using that g_μ is log-concave in the entire positive orthant. The proof also generalizes to sector stable polynomials and more generally it implies spectral independence with respect to all external fields implies entropy contraction.

Lastly, this lemma implies that the MLSI constant for the Glauber dynamics for a k-homogeneous log-concave distribution π is $1/k$. So, the chain mixes in $O(k \log\log n)$ steps.

This was first proved by Cryan-Guo-Mousa. The proof we give here is by Anari-Jain-Koehler-Pham-Vuong.

Proof. First, it turns out that if g_π is a k-homogeneous log-concave function then $f := g_\pi^{1/k}$ is a concave function. We leave this as an exercise. Therefore, by concavity,

$$\forall z_1,\dots,z_n > 0 : f(z_1,\dots,z_k) \le f(1,1,\dots,1) + \sum_{i=1}^n \partial_{z_i} f(1,1\dots,1)\cdot(z_i-1)$$

$$= f(1,\dots,1) + \sum_{i=1}^n \frac{\pi^1(i)}{k}(z_i - 1) \quad ()$$

$$= \sum_{i=1}^n z_i \pi^1(i).$$

The second to last identity uses that g_μ is k-homogeneous, so

$$\partial_{z_i} f(\mathbf{1}) = \frac{1}{k} \partial_{z_i} g_\mu(1, \dots, 1)^{1/k-1} = \pi^1(i).$$

The last identity uses that $f(1, \dot{,} 1) = 1$. ☐

Lemma 1.6 (Entropy Contraction)**.** *Suppose* π *is a log-concave distribution Let* $f : \binom{n}{k} \to \mathbb{R}_{\geq 0}$ *be function. Then,*

$$\mathrm{Ent}_{\pi^1} f^1 \leq \frac{1}{k} \mathrm{Ent}_\pi f$$

Proof. As usual after re-scaling we assume that $\mathbb{E}_\pi f = 1$. So, we can write $f = \mu/\pi$. In such a case,

$$\mathrm{Ent}_\pi f = \mathbb{E} f \log f = \sum_S \mu(S) \log \frac{\mu(S)}{\pi(S)} = D(\mu \| \pi),$$

i.e., it is the KL-divergence of μ, π. Let μ^1 be the marginals of μ. To prove the lemma, we fix the LHS, i.e., $\alpha := \mu^1$ and we look for the distribution which maximizes the RHS, i.e., the KL-divergence. That means we want to find a distribution ν such that $D(\nu \| \pi)$ is maximized subject to $\nu^1 = \alpha$. This exactly corresponds to the convex program in the previous section.

So, by Claim 1.4,

$$\mathrm{Ent}_\pi(f) \geq -\inf_{z>0} \log \frac{g_\pi(z_1, \dots, z_n)}{\prod_i z_i^{\alpha_i}}$$

Now, we are ready to finish the proof: We specialize $z_i \leftarrow \alpha_i/\pi^1(i)$. Then,

$$\begin{aligned}
-\inf_{z>0} \log \frac{g_\pi(z_1, \dots, z_n)}{\prod_i z_i^{\alpha_i}} &\geq -\log \frac{g_\pi(z_1, \dots, z_n)}{\prod_{i=1}^n (\alpha_i/\pi^1(i))^{k\alpha_i}} \\
&\geq -\log \frac{\left(\sum_i \frac{\alpha_i}{\pi^1(i)} \pi^1(i)\right)^{1/k}}{\prod_{i=1}^n (\alpha_i/\pi^1(i))^{k\alpha_i}} && \text{(Lemma 1.5)} \\
&= -\log \prod_{i=1}^n (\pi^1(i)/\alpha_i)^{k\alpha_i} \\
&= \sum_{i=1}^n k\alpha_i \log \frac{\alpha_i}{\pi^1(i)} = k \cdot \mathrm{Ent}_\pi^1 f^1
\end{aligned}$$

as desired. ☐