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

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 \( z^\kappa \) in \( p \), the Newton polytope of \( p \) is the convex hull of all integer vectors \( \kappa \) 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 \( \text{Newt}(p) \) is the spanning tree polytope of \( G \), the convex hull of the indicator vectors of all spanning trees of \( G \).

In this section, we study a generalization of Gurvits’ convex program:
\[ \inf_{z > 0} \frac{p(z_1, \ldots, z_n)}{z^\alpha} \quad (5.1) \]
where \( \alpha \in \mathbb{R}^n_{\geq 0} \).

**Lemma 5.1.** 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 \( \alpha \in \text{Newt}(p) \).

**Proof.** \( \Leftarrow \): First, assume that \( \alpha \in \text{Newt}(p) \). Then, there is a convex combination of the vertices of this polytope that is equal to \( \alpha \),
\[ \alpha = \sum_{\kappa : c_p(\kappa) \neq 0} \lambda_\kappa \kappa \]
where \( \sum_\kappa \lambda_\kappa = 1 \) and each \( \lambda_\kappa \geq 0 \). Then, for any \( z > 0 \) we can write,
\[ p(z) = \sum_{\kappa \in \mathbb{Z}_{\geq 0}^n} \lambda_\kappa \frac{c_p(\kappa) z^\kappa}{\lambda_\kappa} \geq \prod_{\kappa \in \mathbb{Z}_{\geq 0}^n} \left( \frac{c_p(\kappa) z^\kappa}{\lambda_\kappa} \right)^{\lambda_\kappa} = z^\alpha \prod_{\kappa \in \mathbb{Z}_{\geq 0}^n} \left( \frac{c_p(\kappa)}{\lambda_\kappa} \right)^{\lambda_\kappa}, \]
where the inequality follows by the weighted AM-GM inequality and that \( c_p(\kappa) \geq 0 \) and \( z > 0 \). Therefore, \( \inf_{z > 0} \frac{p(z)}{z^\alpha} \geq \prod_{\kappa \in \mathbb{Z}_{\geq 0}^n} \left( \frac{c_p(\kappa)}{\lambda_\kappa} \right)^{\lambda_\kappa} > 0 \) as desired.

\( \Rightarrow \): Conversely, suppose \( \alpha \notin \text{Newt}(p) \). 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 \( x \in \text{Newt}(p) \) 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,
\[ \inf_{z > 0} \frac{p(z)}{z^\alpha} \leq \frac{p(z^*)}{z^{*\alpha}} \]
\[ = \frac{\sum_{\kappa \in \mathbb{Z}_{\geq 0}^n} c_p(\kappa) \exp(\log z^*, \kappa)}{\exp(\log z^*, \alpha)} \]
\[ = \frac{\sum_{\kappa \in \mathbb{Z}_{\geq 0}^n} c_p(\kappa) \exp(t(c, \kappa))}{\exp(t(c, \alpha))} \leq \frac{\sum_{\kappa \in \mathbb{Z}_{\geq 0}^n} c_p(\kappa) \exp(tb)}{\exp(t(b + \epsilon))} \]
Letting $t \to \infty$ the RHS converges to 0.

Some remarks are in order: Recall that in lecture 3 we proved Gurvits’ theorem that for any real stable $p \in \mathbb{R}_{\geq 0}[z_1, \ldots, z_n],
\partial_{z_1} \ldots \partial_{z_n} p |_{z=0} \geq e^{-n} \inf_{z>0} \frac{p(z)}{z_1 \ldots z_n}
The RHS is a special case of (5.1) when $\alpha = 1$. If the RHS is positive, then by the above lemma, $1 \in \text{Newt}(p)$. In such a case Gurvits’ theorem implies that the coefficient of $z^1$ is non-zero in $p$. More generally, this is true for any integer point in Newton polytopes of real stable polynomials: Given any real stable polynomial $p \in \mathbb{R}_{\geq 0}[z_1, \ldots, z_n]$, and any $\alpha \in \mathbb{Z}^n$ such that $\alpha \in \text{Newt}(p)$, we have $c_p(\alpha) > 0.$

Next, we prove the following theorem:

**Theorem 5.2.** Let $\mu : 2^n \to \mathbb{R}_{\geq 0}$ be a probability distribution. Let $\alpha \in \text{Newt}(p)$. 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 $\alpha_i$.

The above theorem conceptually has a very important message. Say $\mu$ is a strongly Rayleigh distribution. It says that given any point $\alpha$ in the Newton polytope of $g_\mu$, there is another strongly Rayleigh distribution $\mu'$ such that the marginals of $\mu'$ is equal to $\alpha$.

**Remark 5.3.** We remark that if $\alpha$ is in the interior of the Newton polytope we can attain $\alpha$ exactly, otherwise, we can only satisfy $\alpha$ 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 $\alpha$.

Recall that many of the probabilistic operations on $\mu$ can be translated to operations on the generating polynomial $g_\mu$. To prove the theorem, it is natural to write down the marginal vector of a distribution $\mu$: For any $1 \leq i \leq n$ we can write
\[ \mathbb{P}_{S*\mu}[i \in S] = \partial_{z_i} g_\mu(z) |_{z=1}. \]

Sometimes, it is cleaner to assume $g_\mu$ is not normalized to $g_\mu(1) = 1$. In such a case, we can write
\[ \mathbb{P}_{S*\mu}[i \in S] = \frac{\sum_{S: i \in S} \mu(S) z^S}{\sum_S \mu(S) z^S} |_{z=1} = z_i \partial_{z_i} \log g_\mu(z) |_{z=1}. \quad (5.2) \]

We write the following convex program and we study its optimality condition.
\[ \inf_y \log \frac{g_\mu(e^{y_1}, \ldots, e^{y_n})}{e^{\langle y, \alpha \rangle}}. \quad \text{(Max-Entropy CP)} \]

Since the above convex program has no constraints, the optimum solution is attained unless the optimum value is $-\infty$. In Lemma 5.1 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
\[ 0 = \partial_{y_i} (\log g_\mu(e^{y_1}, \ldots, e^{y_n}) - \langle y, \alpha \rangle) |_{y=y^*}, \]

Therefore,
\[ \frac{\partial_{y_i} g_\mu(e^{y_1}, \ldots, e^{y_n})|_{y=y^*}}{g_\mu(e^{y^*_1}, \ldots, e^{y^*_n})} = \alpha_i \]
But this means that
\[ \frac{\sum_{S \in S} \mu(S) e^{y^* 1_S}}{\sum_S \mu(S) e^{y^* 1_S}} = \alpha_i \] (5.3)

Letting \( \lambda = e^{y^*} \), by (5.2) we get that
\[ P_{S \sim \lambda \mu}[i] = z_i \partial_{z_i} \log g_{\lambda \mu}(z) |_{z = 1} = \frac{\partial_{z_i} g_{\mu}(\lambda_1 z_1, \ldots, \lambda_n z_n)}{g(\lambda_1, \ldots, \lambda_n)} \]
\[ \sum_{S} \mu(S) \lambda^S = \alpha_i, \]
as desired. The last identity follows by (5.3)

(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 \( \mu \) and evaluate its partial derivatives. If \( \mu \) is a strongly Rayleigh distribution, we can approximately evaluate \( g_{\mu} \). 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 in a few works and it is not in the scope of this course.

5.1 Dual of Max-Entropy CP

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

\[
\begin{align*}
\max & \sum_{\kappa \in \text{Newt}(p)} q_\kappa \log \frac{c_p(\kappa)}{q_\kappa} \\
\text{s.t.,} & \sum_{\kappa \in \text{Newt}(p)} q_\kappa \kappa_i = \alpha_i \quad \forall 1 \leq i \leq n, \\
& \sum_{\kappa} q_\kappa = 1 \\
& q_\kappa \geq 0 \\
\end{align*}
\]

(Max-Entropy Dual)

We claim this is the dual to (Max-Entropy CP). We think of \( q \) as a distribution over integer points in Newt(\( p \)). To write the dual of this program, we first need to write the Lagrangian:

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

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

\[ \max \inf_{q > 0, y \in \mathbb{R}^n, s} L(q, \gamma, s) = \inf_{y \in \mathbb{R}^n, s} \max_{q > 0} L(q, y, s) \] (5.4)

At optimality the gradient of the Lagrangian is zero, so for any \( \kappa \),

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

Therefore, at optimality

\[ \frac{c_p(\kappa)}{q_\kappa} = e^{1-\langle y, \kappa \rangle - s}. \]
Plugging this into (5.4), we can write the dual as follows:

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

(5.5)

\[
= \inf_{y,s} \sum_{\kappa \in \text{Newt}(p)} q_{\kappa} - \langle y, \kappa \rangle - s
\]

(5.6)

\[
= \inf_{y,s} \sum_{\kappa \in \text{Newt}(p)} c_{p}(\kappa) e^{s+y\langle y,\kappa \rangle - 1} - \langle y, \kappa \rangle - s
\]

(5.7)

Optimizing the RHS over \(s\) we get

\[
1 = \sum_{\kappa \in \text{Newt}(p)} c_{p}(\kappa) e^{s+y\langle y,\kappa \rangle - 1} \iff s = -\log \sum_{\kappa \in \text{Newt}(p)} c_{p}(\kappa) e^{y\langle y,\kappa \rangle - 1}
\]

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

\[
\inf_{y} 1 - \langle y, \alpha \rangle + \log \sum_{\kappa \in \text{Newt}(p)} c_{p}(\kappa) e^{y\langle y,\kappa \rangle - 1} = \inf_{y} \log \frac{p(y_1, \ldots, y_n)}{y^n}
\]

as desired.

### 5.2 Applications to TSP

Recall that in the TSP we are given \(n\) cities \(\{1, \ldots, n\}\) and their symmetric pairwise distances, \(c : [n] \times [n] \to \mathbb{R}_+\), we want to find the shortest tour that visits each vertex at least once. Let \(x\) be an optimal solution to the LP relaxation of TSP

\[
\max \sum_{i,j} c(i,j) x_{i,j},
\]

s.t., \(\sum_{i \in S, j \notin S} x_{i,j} \geq 2 \forall S \subseteq V\)

\[
\sum_{j} x_{i,j} = 2 \forall i,
\]

\[
x_{i,j} \geq 0 \forall i, j.
\]

(5.8)

We let \(E\) be the support set of \(x\), i.e., set \(\{i,j\}\) where \(x_{i,j} > 0\) and let \(G = (V, E)\). It turns out that without loss of generality we can assume that there exists an edge \(e^* \in E\) such that \(x_{e^*} = 1\). We define a vector

\[
\alpha = \begin{cases} 
 x_e & \text{if } e \in E \text{ and } e \neq e^* \\
 0 & \text{otherwise}
\end{cases}
\]

It turns out that \(\alpha\) is in the spanning tree polytope of \(G\). Say \(e^* = \{n-1, n\}\). Note that every vertex has fractional degree 2 in \(\alpha\), i.e., for any \(i < n-1\), \(\sum_{e \neq e^*} \alpha_e = 2\).

Let \(\mu\) be the uniform distribution over spanning trees of \((V, E \setminus e^*)\). By Theorem 5.2, there exists an external field \(\lambda\) such that marginals of \(\mu * \lambda\) is equal to \(\alpha\). We use the following algorithm to approximate TSP: We sample \(T \sim \mu * \lambda\); then we add the edge \(e^*\); finally, we add the minimum cost matching on odd degree vertices of \(T \cup \{e^*\}\). It is conjectured that this algorithm gives a better than \(3/2\) approximation for TSP. We are still far from analyzing this algorithm. Here, I show how to use properties of real stable polynomials and SR distributions to prove nice properties of \(T\).
Lemma 5.4. Let \( v_1, \ldots, v_k \) be vertices of \( G \) that do not include \( n - 1, n \) such that the induced graph \( G[\{v_1, \ldots, v_k\}] \) has no edges. Then,

\[
\mathbb{P}_{T \sim \mu \ast \lambda} [d_T(v_1) = \cdots = d_T(v_k) = 2] \geq e^{-k}
\]

where \( d_T(v) \) is the degree of a vertex \( v \) in the sampled tree \( T \).

Proof. Let \( S_1, \ldots, S_k \) be the set of edges incident to \( v_1, \ldots, v_k \) respectively and let \( F \) be the rest of the edges. Note that since \( v_1, \ldots, v_k \) do not share edges, \( S_1, \ldots, S_k \) are mutually disjoint. Define

\[
p(y_1, \ldots, y_k) = g_{\mu \ast \lambda} \left( \begin{array}{c}
z_e = y_1 & \forall e \in S_1, \\
\cdots \\
z_e = y_k & \forall e \in S_k \\
z_e = 1 & \text{otherwise}
\end{array} \right).
\]

Note that in this definition we crucially use that \( S_1, \ldots, S_k \) are disjoint. By closure properties of real stable polynomials \( p \) is real stable. We can re-write \( p \) as follows:

\[
p(y_1, \ldots, y_k) = \sum_T \mu \ast \lambda(T) \prod_{i=1}^k y_i^{d_T(v_i)}.
\]

It follows that

\[
\mathbb{P} [d_T(v_1) = \cdots = d_T(v_k) = 2] = 2^k \partial_{y_1}^2 \cdots \partial_{y_k}^2 p \big|_{y=0},
\]

i.e., the RHS is the coefficient of \( y_1^2 \cdots y_k^2 \) in \( p \). Furthermore, note that each of the vertices \( v_1, \ldots, v_k \) have degree at least 1 in \( T \); so we can factor out a monomial \( y_1 \cdots y_k \),

\[
p(y) = y_1 \cdots y_k q(y_1, \ldots, y_k).
\]

It follows that \( q \) is also real stable. So, we need to show that

\[
\partial_{y_1} \cdots \partial_{y_k} q \big|_{y=0} \geq e^{-k}.
\]

Since \( q \) is real stable and has non-negative coefficients, by Theorem 3.1, we have

\[
\partial_{y_1} \cdots \partial_{y_k} q \big|_{y=0} \geq e^{-k} \inf_{y>0} \frac{q(y)}{y_1 \cdots y_k}.
\]

So, all we need to show is that

\[
\inf_{y>0} \frac{q(y)}{y_1 \cdots y_k} \geq 1. \tag{5.9}
\]

First, observe that we can write \( q \) as follows:

\[
q(y_1, \ldots, y_k) = \sum_T \mu \ast \lambda(T) \prod_{i=1}^k y_i^{d_T(v_i)-1}
\]

\[
\geq \prod_T \left( \prod_{i=1}^k y_i^{d_T(v_i)-1} \right)^{\mu \ast \lambda(T)}
\]

\[
= \prod_{i=1}^k y_i^{\sum_T \mu \ast \lambda(T) d_T(v_i)-1} = \prod_{i=1}^k y_i^{2-1},
\]

where the inequality follows by weighted AM-GM and the last identity follows by the fact that \( \mathbb{E} [d_T(v)] = 2 \) for any vertex other than \( n - 1, n \). This proves (5.9). \( \square \)
The following generalization is proved in my work with Karlin and Klein:

**Theorem 5.5.** Given a SR distribution \( \mu : 2^{[n]} \to \mathbb{R}_+ \), and disjoint sets \( A_1, \ldots, A_k \) and integers \( n_1, \ldots, n_k \) such that for any \( S \subseteq [k] \),

\[
P_{T \sim \mu} \left[ |T \cap \bigcup_{i \in S} A_i| = \sum_{i \in S} n_i \right] \geq \epsilon
\]

Then,

\[
P \left[ \forall i : |T \cap A_i| = n_i \right] \geq \epsilon^{2^k f(n_1, \ldots, n_m)}.
\]

The above bound is not ideal; in particular, we expect only an exponential dependency on \( k \) in the RHS such as \( \epsilon^{O(k)} \).