Lecture 12: Polynomial Capacity

Recall: Geom. of polys. method:

(1) Encode object you care about as nice polynomial (stable, Lorentzian, ..)

(2) Apply operators which preserve nice properties

(3) Extract info relating back to original object

Before: info was log-concavity of coefficients/function, or Rayleigh condition

Theorem: For any Lorentzian polyn. p of deg. d, and any i, j,

$$2\left(1-\frac{1}{d}\right)\partial_{x_i} p \cdot \partial_{x_j} p - p \cdot \partial_{x_i}\partial_{x_j} p \geq 0$$

on $\mathbb{R}^n_{\geq 0}$.

(Note for $d=2$, this is the strong Rayleigh condition on $\mathbb{R}^n_{\geq 0}$.)

(Proof in HW.)

New info. we want to study: coefficients/evaluations/inner products of polynomials.

<u>First motivation</u>: Permanent of a matrix / perfect matchings of a bipartite graph.

<u>Def.</u>: Given a matrix A, its <u>permanent</u> is given by

$$\mathrm{per}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i,\sigma(i)}.$$

"Like the determinant, only simpler."

↳ poly-time comp. for det, given by Gaussian elimination

↳ #P-hard to compute per exactly, even for 0-1 matrix

(Note: can <u>search</u> for perf. matching in poly-time, but <u>counting</u> is hard)

(How does this connect to polynomials?)

Given $A \in \mathbb{R}_{\geq 0}^{n \times n}$, define

$$p_A(x) = \prod_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_j \in \mathbb{R}_{\geq 0}[x_1, \dots, x_n]$$

is n-homog. and real stable (which implies Lorentzian; HW).

I.e.:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots \\ a_{21} & a_{22} & a_{23} & \cdots \\ a_{31} & a_{32} & a_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \Longrightarrow p_A(x) = \begin{array}{l} (a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots) \\ \cdot (a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots) \\ \cdot (a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \dots) \\ \cdot \cdots \end{array}$$

Expand $p_A \Rightarrow$ possible $x_1 x_2 \cdots x_n$ terms correspond precisely to $\prod_{i=1}^{n} a_{i,\sigma(i)}$ for each $\sigma \in S_n$.

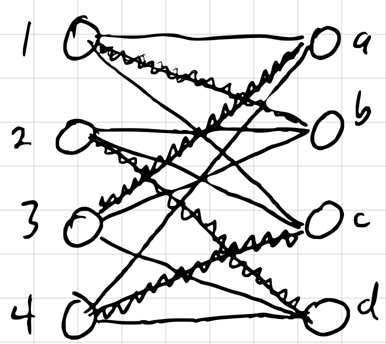
Thus, $\langle x^{\mathbb{1}} \rangle p_A = \mathrm{per}(A)$.

I.e., $\mathrm{per}(A) = \partial_{x_1}|_{x_1=0} \cdots \partial_{x_n}|_{x_n=0}\, p$.

(This gives us a way to induct on degree/# vars, but not in such a simple way as w/ Matroid polynomials.)

<u>Bipartite graph version</u>:

Let G be a bipartite graph on $n+n$ vertices. E.g.:



$\Longleftrightarrow$

bipartite adj. matrix A

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

perf. matchings $=$ per(A)

d-regular $\Longleftrightarrow$ row sums $=$ col sums $= d$

$$\#pm(G) = \text{per}(A) = \partial_{x_1}|_{x_1=0} \cdots \partial_{x_n}|_{x_n=0}\, p$$

(existence of p.m. $\Longleftrightarrow$ per >0)

Van der Waerden "conjecture" ('30s):

If row sums / col sums $A \in \mathbb{R}_{\geq 0}^{n \times n}$ are all equal to 1 ("doubly stochastic")

then: $per(A) \geq \frac{n!}{n^n}$.

(Proven by Falikman, Egorychev '81; original proofs used AF ineq. and were complicated. Easier?)

Cor. If G is d-regular bipartite, then $\#p.m.(G) \geq d^n \cdot \frac{n!}{n^n}$.

Cor.: 1 is an e^n-approximation to per(A) for any D.S. A.

Pf: $1 \geq per(A) \geq \frac{n!}{n^n} \geq e^{-n}$

Stirling's approx.

$\langle x^{\mathbb{1}} \rangle_{p_A} \leq p_A(\mathbb{1}) \leq 1$.

Cor.: Approx. algo. to per(A) for any $A \in \mathbb{R}_{\geq 0}^{n \times n}$.

Pf.: Sinkhorn Scaling algo. iteratively converts A into a D.S. matrix, with easy-to-track changes to per.

<u>Proof Idea</u>: Generalize conjecture to Lorentzian polynomials.

Notice, $p_A(\mathbb{1}) = \prod_{i=1}^{n} \sum_{j=1}^{n} a_{ij} = 1$.

Also, $(\partial_{x_\ell} p_A)(\mathbb{1}) = \sum_{k=1}^{n} a_{k\ell} \cdot \prod_{i \neq k} \sum_{j=1}^{n} a_{ij} = 1$.

(product rule)

Thus, $p_A(\mathbb{1}) = 1$, $\nabla p_A(\mathbb{1}) = \mathbb{1}$.

Call such a polynomial <u>doubly stochastic</u>.

<u>New conj.</u>: The all-ones coeff of a Lorentzian D.S. polynomial is at least $\frac{n!}{n^n}$.

(Note that d-homog. $\Rightarrow$

$$d \cdot p(\mathbb{1}) = \sum_{i=1}^{n} (\partial_{x_i} p)(\mathbb{1}) = \mathbb{1} \cdot \nabla p_A(\mathbb{1}),$$

so that $d = n$ in this case.)

(We will prove this using <u>Gurvits' theorem</u>. First need notion of <u>polynomial capacity</u>.)

Defn.: Given $p \in \mathbb{R}_{\geq 0}[x_1, \dots, x_n]$ and $\alpha \in \mathbb{R}^n_{\geq 0}$, define:

$$\mathrm{Cap}_\alpha(p) := \inf_{\substack{x > 0 \\ (x \in \mathbb{R}^n_{>0})}} \frac{p(x)}{x^\alpha}.$$

(We will discuss more properties of this quantity later.)

Theorem (Gurvits '05, '09): If $p \in \mathbb{R}_{\geq 0}[x_1, \dots, x_n]$ is d-homog. and Lorentzian, then $\forall \mu \in \mathbb{Z}^n_{\geq 0}$ with $|\mu| = d$,

$$\langle x^\mu \rangle p \geq \binom{d}{\mu} \cdot \frac{\mu^\mu}{d^d} \cdot \mathrm{Cap}_\mu p.$$

(proof next time)

(How to use to prove Conj.?)

Cor.: For doubly stochastic Lorentzian p, we have

$$\langle x^{\mathbb{1}} \rangle p \underset{\substack{\uparrow \\ \text{Gurvits' thm.}}}{\geq} \frac{n!}{n^n} \cdot \mathrm{Cap}_{\mathbb{1}}(p) \underset{\substack{\uparrow \\ \text{How?}}}{=} \frac{n!}{n^n}.$$

Lemma: Let $p \in \mathbb{R}_{\geq 0}[x_1, \dots, x_n]$ be s.t. $p(\bar{1}) = 1$ and $\nabla p(\bar{1}) = \alpha$. Then $\mathrm{Cap}_\alpha(p) = 1$.

Proof: Note that if $p(x) = \sum_{\kappa \in \mathbb{Z}^n_{\geq 0}} p_\kappa x^\kappa$, then $\alpha = \nabla p(\bar{1}) = \sum_{\kappa \in \mathbb{Z}^n_{\geq 0}} p_\kappa \cdot \kappa$

Since $1 = p(\bar{1}) = \sum_{\kappa \in \mathbb{Z}^n_{\geq 0}} p_\kappa$,

α is the expectation of the probability distribution μ on $\mathbb{Z}^n_{\geq 0}$ given by $\mathbb{P}[\mu = \kappa] = p_\kappa$.

By the weighted AM-GM ineq. (side board; e.g. via Jensen's ineq.), we have:

$$\sum_\kappa p_\kappa x^\kappa \geq \prod_\kappa (x^\kappa)^{p_\kappa} = x^{\sum_\kappa p_\kappa \cdot \kappa} = x^\alpha$$

for all $x \in \mathbb{R}^n_{>0}$.

Thus $\dfrac{\sum_\kappa p_\kappa x^\kappa}{x^\alpha} \geq 1 \quad \forall x \in \mathbb{R}^n_{>0}$.

Further, $\frac{\rho(1)}{(1)^{\alpha}} = 1 \Rightarrow$

$$\mathrm{Cap}_{\alpha}(\rho) = \inf_{x>0} \frac{\rho(x)}{x^{\alpha}} = 1. \quad \square$$

(Proof hints at bigger picture for the interpretation of capacity.)

Lecture 13: Polynomial Capacity and Gurvits Theorem

Last time: Given $A \in \mathbb{R}_{\geq 0}^{n \times n}$, construct

$$p_A(x) = \prod_{i=1}^{n} (Ax)_i = \prod_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_j .$$

Then $\langle x^{\mathbb{1}} \rangle p_A(x) = \mathrm{per}(A)$.
(If A is bipartite adjacency matrix, then this also counts perfect matchings.)

Gurvits' Theorem: If $p \in \mathbb{R}_{\geq 0}[x_1, \ldots, x_n]$ is d-homog. and Lorentzian, then for all $\mu \in \mathbb{Z}_{\geq 0}^n$, $|\mu| = d$,

$$\mathrm{Cap}_\mu(p) \geq \langle x^\mu \rangle p(x) \geq \binom{d}{\mu} \frac{\mu^\mu}{d^d} \cdot \mathrm{Cap}_\mu(p).$$

Stirling

$$\geq \sqrt{2\pi d} \cdot \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \mu_i}} \cdot \mathrm{Cap}_\mu(p).$$

Cor: For D.S. $A \in \mathbb{R}_{\geq 0}^{n \times n}$, $1 \geq \mathrm{per}(A) \geq \frac{n!}{n^n}$

Also recall:

Lemma: Let $p \in \mathbb{R}_{\geq 0}[x_1, \dots, x_n]$ be s.t. $p(\bar{1}) = 1$ and $\nabla p(\bar{1}) = \alpha$.
Then $\mathrm{Cap}_\alpha(p) = 1$.

Proof used following fact:
Since $1 = p(\bar{1}) = \sum_{\kappa \in \mathbb{Z}^n_{\geq 0}} p_\kappa$,

α is the expectation of the probability distribution μ on $\mathbb{Z}^n_{\geq 0}$ given by $\mathbb{P}[\mu = \kappa] = p_\kappa$.
That is, $\nabla p(\bar{1})$ is the vector of marginal (probabilities) of μ (when p is multiaffine).

$$\left(\mathrm{Cap}_\alpha(p) = \inf_{x > 0} \frac{p(x)}{x^\alpha}.\right)$$

Other properties of Capacity:

For $p \in \mathbb{R}_{\geq 0}[x_1, \dots, x_n]$, $p(\mathbb{1}) = 1$,

(1) $\mathrm{Cap}_\alpha(p) \leq 1$, $\forall \alpha \in \mathbb{R}^n_{\geq 0}$

(2) If $\mu \in \mathrm{Supp}(p)$, then $p_\mu \leq \mathrm{Cap}_\mu(p)$.

(3) $\mathrm{Cap}_\alpha(p) > 0 \Leftrightarrow \alpha \in \mathrm{Newt}(p)$

↳ Newton polytope of p

(4) $\log \mathrm{Cap}_\alpha(p)$ is a convex program via $x \mapsto e^y$, so that it is efficiently computable:

Fenchel conjugate

$$\log \mathrm{Cap}_\alpha(p) = \inf_{y \in \mathbb{R}^n} \left[-\langle y, \alpha \rangle + \log \sum_\kappa p_\kappa e^{\langle y, \kappa \rangle} \right]$$

(5) $x^* = \arg \mathrm{Cap}_\alpha(p) \Rightarrow$ coeff. of $\dfrac{p(x^* \odot x)}{p(x^*)}$ give the "entropy maximizing distr." on $\mathrm{supp}(p)$ with expectation α.

(Notice we have

$$\log \mathrm{Cap}_\alpha(p) = \inf_{y \in \mathbb{R}^n} \log \sum_\kappa p_\kappa e^{\langle y, \kappa - \alpha \rangle},$$

and $\nabla|_{y=0}(\text{obj.}) = \dfrac{\sum_\kappa p_\kappa (\kappa - \alpha)}{\sum_\kappa p_\kappa} = \nabla p(\mathbb{1}) - \alpha$.)

Proof: (1) Plug in $x = \mathbb{1}$.

(2) $\mathrm{Cap}_\mu(p) = \inf_{x>0} \frac{p(x)}{x^\mu} \geq \inf_{x>0} \frac{p_\mu x^\mu}{x^\mu} = p_\mu$

(3) ($\Leftarrow$) Let $\alpha = \sum_{\kappa \in \mathrm{supp}(p)} c_\kappa \cdot \kappa$, convex combo,

and let $\gamma = \min_{\kappa \in \mathrm{supp}(p)} \frac{p_\kappa}{c_\kappa} > 0$.

By weighted AM-GM ineq.,

$$p(x) = \sum_\kappa p_\kappa x^\kappa = \sum_\kappa \frac{p_\kappa}{c_\kappa} c_\kappa x^\kappa \geq \gamma \cdot \sum_\kappa c_\kappa x^\kappa$$

$$\geq \gamma \cdot \prod_\kappa (x^\kappa)^{c_\kappa} = \gamma \cdot x^{\sum_\kappa c_\kappa \cdot \kappa} = \gamma \cdot x^\alpha,$$

for all $x > 0$. Thus,

$$\mathrm{Cap}_\alpha(p) = \inf_{x>0} \frac{p(x)}{x^\alpha} \geq \inf_{x>0} \frac{\gamma \cdot x^\alpha}{x^\alpha} = \gamma > 0.$$

($\Rightarrow$) Contrapositive. Suppose $\alpha \notin \mathrm{Newt}(p)$.

Then $\exists$ separating hyperplane, i.e.,

$\exists \beta \in \mathbb{R}^n$ s.t. $\langle \beta, \kappa - \alpha \rangle < 0$

$\forall \kappa \in \mathrm{supp}(p)$. By (4), $\log \mathrm{Cap}_\alpha(p)$

$$\leq \inf_{t \in \mathbb{R}} \left[-\langle t\beta, \alpha \rangle + \log \sum_\kappa p_\kappa e^{\langle t\beta, \kappa \rangle} \right]$$

$$= \inf_{t \in \mathbb{R}} \log \sum_\kappa p_\kappa e^{t \langle \beta, \kappa - \alpha \rangle} = -\infty.$$

Thus, $\mathrm{Cap}_\alpha(p) = 0$.

Proof of Gurvits' theorem

(Need a few Lemmas.)

Lemma (BLP '20): Let $q, w \in \mathbb{R}^d_{\geq 0}[t]$ be such that w has all pos. coeff. and $\left(\frac{q_k}{w_k}\right)_{k=0}^d$ forms a log-concave sequence (with no holes). Then for all $0 \leq k \leq d$, we have

$$q_k \geq \frac{w_k}{\mathrm{Cap}_k(w)} \cdot \mathrm{Cap}_k(q).$$

Proof: Define $a_k = \frac{q_k}{w_k}$. Equiv.:

$$\mathrm{Cap}_k(w) \geq \frac{\mathrm{Cap}_k(q)}{a_k} = \inf_{t>0} \sum_{i=0}^{d} \frac{a_i w_i}{a_k} t^{i-k}.$$

WLOG, we may assume $a_k = 1 \Rightarrow$ (divide a_i seq. by a_k)

To prove: $\mathrm{Cap}_k(w) \geq \inf_{t>0} \sum_{i=0}^{d} a_i w_i t^{i-k}$.

Log-concavity of (a_i): $a_i^2 \geq a_{i-1} a_{i+1}$

$$\Rightarrow a_{i+1} \leq \frac{a_i^2}{a_{i-1}}, \quad a_{i-1} \leq \frac{a_i^2}{a_{i+1}},$$

$$\frac{a_{i+1}}{a_i} \leq \frac{a_i}{a_{i-1}}, \quad \frac{a_{i-1}}{a_i} \leq \frac{a_i}{a_{i+1}}.$$

Now, $a_k = 1$ by assumption.

<u>Claim</u>: $a_{k+i} \le a_{k+1}^i \quad \forall i \in \mathbb{Z}$

(if $a_{k+1} \ne 0$; proof easier if $a_{k+1} = 0$)

By two-sided induction.

($i = 0, 1 \to$ trivial.) $\quad i > 1 \to$

$$a_{k+i} \le \frac{a_{k+i-1}^2}{a_{k+i-2}} \le a_{k+i-1}\left(\frac{a_{k+1}}{a_k}\right)$$

$$\le a_{k+1}^{i-1} \cdot a_{k+1} = a_{k+1}^i .$$

$i < 0 \to$

$$a_{k+i} \le \frac{a_{k+i+1}^2}{a_{k+i+2}} \le a_{k+i+1} \cdot \left(\frac{a_k}{a_{k+1}}\right)$$

$$\le a_{k+1}^{i+1} \cdot a_{k+1}^{-1} = a_{k+1}^i .$$

(Really this is just log-concavity of $(a_i)_i$ as a discrete function.)

Thus, $\inf\limits_{t>0} \sum\limits_{i=0}^{d} a_i w_i t^{i-k}$

$$= \inf_{t>0} \sum_{i=-k}^{d-k} w_{k+i}\, a_{k+i}\, t^i$$

$$\le \inf_{t>0} \sum_{i=-k}^{d-k} w_{k+i} \left(a_{k+1} t\right)^i$$

$$= \inf_{t>0} \sum_{i=-k}^{d-k} w_{k+i} t^i$$

$$= \inf_{t>0} \sum_{i=0}^{d} w_i t^{i-k}$$

$$= Cap_k(w).$$

If $a_{k+1} = 0$, then no holes $\Rightarrow$

$$\inf_{t>0} \sum_{i=0}^{d} a_i w_i t^{i-k}$$

$$= \inf_{t>0} \left[\sum_{i=0}^{k-1} a_i w_i t^{i-k} + w_k \right]$$

$$= w_k + \inf_{t>0} \sum_{i=-k}^{-1} a_{k+i} w_{k+i} t^i$$

$$= w_k \leq Cap_k(w). \qquad \square$$

<u>Lemma</u>: If $w(t) = (t+1)^d$, then $Cap_k(w) = \frac{d^d}{k^k (d-k)^{d-k}}$.

(<u>Note</u>: $w(t) = (t+1)^d \Rightarrow$ q_k / w_k log-concave $\Leftrightarrow q_k$ ultra log-concave.)

Proof: $\mathrm{Cap}_k(\omega) = \inf_{t>0} \frac{(1+t)^d}{t^k}$

$$= \left[\inf_{t>0} \frac{1+t}{t^{k/d}}\right]^d = \left[\inf_{t>0}\left(t^{-k/d} + t^{1-k/d}\right)\right]^d$$

$$0 = \partial_t\left[t^{-k/d} + t^{1-k/d}\right] = -\frac{k}{d} t^{-1-\frac{k}{d}} + \left(1-\frac{k}{d}\right) t^{-\frac{k}{d}}$$

$$= t^{-1-\frac{k}{d}}\left(-\frac{k}{d} + \left(1-\frac{k}{d}\right)t\right)$$

$$\Rightarrow t = \frac{k/d}{1-k/d} = \frac{k}{d-k}$$

$$\Rightarrow \mathrm{Cap}_k(\omega) = \frac{\left(1+\frac{k}{d-k}\right)^d}{\left(\frac{k}{d-k}\right)^k} = \frac{d^d}{k^k (d-k)^{d-k}}. \quad \square$$

Gurvits' Theorem: If $p \in \mathbb{R}_{\geq 0}(x_1, \dots, x_n)$ is d-homog. and Lorentzian, then for all $\mu \in \mathbb{Z}^n_{\geq 0}$, $|\mu| = d$,

$$\mathrm{Cap}_\mu(p) \geq \langle x^\mu \rangle p(x) \geq \binom{d}{\mu} \frac{\mu^\mu}{d^d} \cdot \mathrm{Cap}_\mu(p).$$

Proof: Induction on n. Trivial for $n=1$ ($\mu = (d)$).

Now consider

$$\inf_{x>0} \frac{p(x)}{x^\mu} = \inf_{x_1>0} \cdots \inf_{x_n>0} \frac{p(x_1, \dots, x_n)}{x_1^{\mu_1} \cdots x_n^{\mu_n}}$$

Fix $x_1, \dots, x_{n-1} > 0$. Then:

$$\inf_{x_n > 0} \frac{p(x_1, \ldots, x_{n-1}, x_n)}{x_n^{\mu_n}}$$

$$= \inf_{t > 0} \frac{p(x_1, \ldots, x_{n-1}, t)}{t^{\mu_n}} = \mathrm{Cap}_{\mu_n}(q)$$

where $q(t) := p(x_1, \ldots, x_{n-1}, t)$

Defining $f(t, s) := p(x_1 s, \ldots, x_{n-1} s, t)$

$f(s, t)$ is Lorentzian degree d, so the coeff. of q are ultra log-concave (w.r.t. degree d).

Thus by the lemmas using $w(t) = (t+1)^d$, we have

$$q_{\mu_n} \geq \frac{w_{\mu_n}}{\mathrm{Cap}_{\mu_n}(w)} \cdot \mathrm{Cap}_{\mu_n}(q)$$

$$= \frac{d!}{\mu_n!(d-\mu_n)!} \cdot \frac{\mu_n^{\mu_n}(d-\mu_n)^{d-\mu_n}}{d^d}$$

$$\cdot \inf_{x_n > 0} \frac{p(x_1, \ldots, x_{n-1}, x_n)}{x_n^{\mu_n}}.$$

Divide by $x_i^{\mu_i}$ $(1 \le i \le n-1)$ and take infs to get:

$$\inf_{x_1,\dots,x_{n-1}>0} \frac{q_{\mu_n}(x_1,\dots,x_{n-1})}{x_1^{\mu_1}\cdots x_n^{\mu_n}} \geq \overbrace{\left(\frac{d!}{\mu_n!(d-\mu_n)!}\cdot\frac{\mu_n^{\mu_n}(d-\mu_n)^{d-\mu_n}}{d^d}\right)}^{C_{\mu_n}} \mathrm{Cap}_\mu(p).$$

Now, $q_{\mu_n}(x_1,\dots,x_{n-1}) = \frac{1}{\mu_n!}\partial_{x_n}^{\mu_n}\big|_{x_n=0} p(x)$

$$\Rightarrow \mathrm{Cap}_{(\mu_1,\dots,\mu_{n-1})}\left(\frac{1}{\mu_n!}\partial_{x_n}^{\mu_n}\big|_{x_n=0} p(x)\right) \geq C_{\mu_n}\cdot \mathrm{Cap}_\mu(p)$$

(Idea of "Capacity preserving operators": derivative operator can only decrease capacity by so much)

Now, by induction,

$$p_\mu = \langle x_1^{\mu_1} \cdots x_{n-1}^{\mu_{n-1}} \rangle \frac{1}{\mu_n!} \partial_{x_n}^{\mu_n}\Big|_{x_n=0} p(x)$$

$$\geq \binom{d-\mu_n}{\mu_1, \ldots, \mu_{n-1}} \cdot \frac{\mu_1^{\mu_1} \cdots \mu_{n-1}^{\mu_{n-1}}}{(d-\mu_n)^{d-\mu_n}}$$

$$\cdot \operatorname{Cap}_{(\mu_1, \ldots, \mu_{n-1})}\left(\frac{1}{\mu_n!} \partial_{x_n}^{\mu_n}\Big|_{x_n=0} p \right),$$

since $\deg\left(\frac{1}{\mu_n!} \partial_{x_n}^{\mu_n}\Big|_{x_n=0} p \right) = d - \mu_n$.

Thus,

$$p_\mu \geq \frac{(d-\mu_n)!}{\mu_1! \cdots \mu_{n-1}!} \cdot \frac{\mu_1^{\mu_1} \cdots \mu_{n-1}^{\mu_{n-1}}}{(d-\mu_n)^{d-\mu_n}}$$

$$\cdot \frac{d!}{\mu_n!(d-\mu_n)!} \cdot \frac{\mu_n^{\mu_n}(d-\mu_n)^{d-\mu_n}}{d^d} \cdot \operatorname{Cap}_\mu(p).$$

(Simplifying gives the result.) $\square$

(Note that if p real stable, $q(t)$ has ultra log-concave coeff. w.r.t. $\deg(q(t)) \Rightarrow$ better bounds via per-variable deg.)