

# Capacity Recap: All the forms of capacity

## Polynomial Capacity: Theory, Applications, Generalizations

Jonathan Leake

Technische Universität Berlin

February 25th, 2021

- 1 Last time: Invariant-theoretic capacity
  - Null-cone problem
  - Invariant-theoretic capacity and the scaling-type algorithm
  - Scaling for real stable polynomials
- 2 Remembering all the forms of capacity
  - Polynomial capacity
  - Entropic capacity
  - Matrix capacity
  - Invariant-theoretic (non-commutative) capacity
- 3 Thanks!

- 1 Last time: Invariant-theoretic capacity
  - Null-cone problem
  - Invariant-theoretic capacity and the scaling-type algorithm
  - Scaling for real stable polynomials
- 2 Remembering all the forms of capacity
  - Polynomial capacity
  - Entropic capacity
  - Matrix capacity
  - Invariant-theoretic (non-commutative) capacity
- 3 Thanks!

# The null-cone problem

Let  $\pi : G \rightarrow \text{GL}(V)$  be a representation of a group  $G$  (i.e.,  $\pi$  is a group homomorphism and  $V$  is a vector space).

**Definition:** An **orbit** of  $v \in V$  is the set  $\mathcal{O}_v := \{\pi(g)v : g \in G\} \subset V$ .

**Definition:** The **null-cone** of  $V$  or  $\pi$  is the set  $\{v : 0 \in \overline{\mathcal{O}_v}\}$ .

**[Hilbert], [Mumford '65]:**  $v$  is in the null-cone iff for every non-constant homogeneous  $G$ -invariant polynomial  $p$  on  $V$  we have  $p(v) = 0$ .

**E.g.:**  $v$  in null-cone  $\implies \pi(g_i)v \rightarrow 0 \implies p(v) = p(\pi(g_i)v) = p(0) = 0$ .

**[Kempf-Ness '79]:**  $v$  is *not* in the null-cone iff  $\mu(w) = 0$  for some  $w \in \overline{\mathcal{O}_v}$ , where  $\mu$  is the **moment map** of  $\pi$ .

**Moment map:** Something like the “gradient” of the action of  $\pi$  at  $g = \text{id}$ :

$$“\mu(w) = \nabla|_{X=0} \log \|\pi(e^X)w\|”.$$

**Convex programming:**  $f = \|w\|$  attains minimum at  $w_0$  iff  $\nabla f(w_0) = 0$ .

# The moment map and moment polytope

**Throughout:** Think  $G = \mathrm{GL}_n(\mathbb{C})$  or  $G = \mathbb{T}^n$  with  $\pi : G \rightarrow \mathrm{GL}(V)$ .

**Definition:** The moment map  $\mu(v)$  for  $v \in V$  is defined via

$$\langle H, \mu(v) \rangle := \partial_t|_{t=0} \log \|\pi(e^{tH})v\|,$$

and  $\mu(v)$  is Hermitian for  $\mathrm{GL}_n(\mathbb{C})$  or a real (diagonal) vector for  $\mathbb{T}^n$ .

**Idea:**  $\mu(v)$  is the “gradient” of  $\log \|\pi(e^X)v\|$  at  $X = 0$ .

**Moment polytope:**  $\Delta(v) := \overline{\{\mathrm{eig}(\mu(w)) : w \in \mathcal{O}_v\}}$  is a convex polytope.

**Kempf-Ness:**  $v$  not in null-cone iff  $\mu(w) = 0$  for a  $w \in \overline{\mathcal{O}_v}$  iff  $\mathbf{0} \in \Delta(v)$ .

**Recall:** In the commutative case ( $G = \mathbb{T}^n$ ), we have:

- $\inf_{g \in G} \|\pi(g) \cdot v\|_2^2 = \inf_{\mathbf{x} > 0} \sum_k |c_k|^2 \mathbf{x}^{2\omega_k}$  is a capacity problem.
- **Kempf-Ness:**  $\mathrm{Cap}_0(\sum_k |c_k|^2 \mathbf{x}^{2\omega_k}) > 0$  iff  $\mathbf{0} \in \mathrm{Newt}(\sum_k |c_k|^2 \mathbf{x}^{2\omega_k})$ .
- Roughly equivalent to polynomial capacity.

# Invariant-theoretic capacity

**Last slide:**  $\inf_g \|\pi(g)v\|$  is a capacity problem in the commutative case.

In more general cases, let's just **make this the definition**:

$$\text{Cap}_0(v) := \inf_{g \in G} \|\pi(g)v\|.$$

“Non-commutative” capacity, “invariant-theoretic” capacity, etc.

Also called **non-commutative geometric programming** since the commutative case captures unconstrained **geometric programming** (see [Bürgisser-Li-Nieuwboer-Walter '20]).

**Kempf-Ness:**  $\text{Cap}_0(v) > 0$  iff  $\mathbf{0}$  is in the moment polytope  $\implies$  Generalization of the same statement for polynomial capacity.

**Recall:**  $\inf_{\mathbf{y} \in \mathbb{R}^n} \log \sum_{k=1}^n |c_k|^2 e^{\langle \mathbf{y}, 2\omega_k \rangle}$  is a convex program. **Can we do the same thing to non-commutative capacity?**

Appears to be “no”... but general capacity is still **geodesically convex**.  
**And, there is a scaling-type algorithm.**

# Scaling-type algorithm

**Recall** the scaling-type algo, where the measure of progress is  $\mu = \text{Cap}_0$ :

- 1 **“Preprocess”**: Set  $g_0 = \text{id}$ .
- 2 **Iterations**: Geodesic gradient descent, Taylor approx, “trust-region” methods... **I.e.**: Natural analogs to convex Euclidean techniques.
- 3 **Approximation**: How close do we need to get before stopping?

**Approximation step is key to determine computational complexity.**

**Theorem [BFGOWW '19]**: For  $\|v\| = 1$ , we have

$$1 - \frac{\|\mu(v)\|}{\gamma(\pi)} \leq [\text{Cap}_0(v)]^2 \leq 1 - \frac{\|\mu(v)\|^2}{4N(\pi)^2}.$$

**Corollary**:  $\mathbf{0} \in \Delta(v)$  iff  $\Delta(v)$  contains a point smaller than  $\gamma(\pi)$ .  
(This  $\gamma(\pi)$  is how close we must get before stopping.)

**Definition**: The **weight margin**  $\gamma(\pi)$  is the minimum distance between  $\mathbf{0}$  and any subset of the “weights” whose convex hull does not contain  $\mathbf{0}$ .

# Weight margin examples

**Last slide:** The **weight margin**  $\gamma(\pi)$  is the minimum distance between  $\mathbf{0}$  and any subset of the “weights” whose convex hull does not contain  $\mathbf{0}$ .

**Matrix scaling:** Action of  $(\mathbb{S}\mathbb{T}^n)^2$  via left-right action on matrices.  
 $\gamma(\pi) \geq \frac{1}{\text{poly}(n)}$  via [Linial-Samorodnitsky-Wigderson '00].

**Operator scaling:** Action of  $(\text{SL}_n(\mathbb{C}))^2$  on  $(M_1, \dots, M_\ell)$  via simultaneous left-right action.  $\gamma(\pi) \geq \frac{1}{\text{poly}(n)}$  via [Gurvits '04], [GGOW '15].

**Tensor scaling for 3-tensors:** Action of  $(\text{GL}_n(\mathbb{C}))^3$  on 3-tensors.  
 $\gamma(\pi) \leq 2^{-\text{poly}(n)}$  via [Franks-Reichenbach '21] (the other day).

**Last result:** Negative result for this method. **Open:** Other methods?

**Real stable polynomials formulation:** Given a real stable polynomial with  $\mathbf{1}$  not in its Newton polytope, how far away can Newton polytope be?



# Weight margin for real stable polynomials

**Theorem [BFGOWW '19]:** For  $\|v\| = 1$ , we have

$$\text{Cap}_0(v) = \inf_{g \in G} \|\pi(g)v\|_2 \geq \left(1 - \frac{\|\mu(v)\|}{\gamma(\pi)}\right)^{1/2}.$$

**Theorem [L-Gurvits '20]:** Fix an  $n$ -homogeneous  $n$ -variate real stable  $p$  with non-negative coefficients. If  $p(\mathbf{1}) = 1$  and  $\|\mathbf{1} - \nabla p(\mathbf{1})\|_1 < 2$ , then

$$\text{Cap}_1(p) = \inf_{x > 0} \frac{p(\mathbf{x})}{\mathbf{x}^1} \geq \left(1 - \frac{\|\mathbf{1} - \nabla p(\mathbf{1})\|_1}{2}\right)^n.$$

**Corollary:** If  $\mathbf{1} \notin \text{Newt}(p)$ , then  $\|\mathbf{1} - \text{Newt}(p)\|_1 \geq 2$ .

**Easy:** (Kempf-Ness)  $\mathbf{1} \notin \text{Newt}(p)$  iff  $\text{Cap}_1(p) = 0$ .

$\implies$  Above bound cannot hold.

$\implies \|\mathbf{1} - \nabla p(\mathbf{1})\|_1 \geq 2$ .

**Finally:** "Scale"  $p$  to have marginals close to the boundary of  $\text{Newt}(p)$ .

**Corollary:** Nice weight margin for special subclass of polynomials. **Can we do something similar in the more general non-commutative case?**

# Scaling for real stable polynomials

**Proof:** Take any  $\alpha$  in the relative interior of  $\text{Newt}(p)$ .

- $\text{Cap}_\alpha(p) > 0$  **and** there is a  $\mathbf{y} > 0$  such that  $\frac{p(\mathbf{y})}{\mathbf{y}^\alpha} = \text{Cap}_\alpha(p)$ .
- Defining  $q(x) := \frac{p(\mathbf{y} \circ x)}{p(\mathbf{y})}$ , we have

$$\text{Cap}_\alpha(q) = \inf_{x>0} \frac{p(\mathbf{y} \circ x)}{x^\alpha \cdot p(\mathbf{y})} = \frac{\mathbf{y}^\alpha}{p(\mathbf{y})} \cdot \inf_{x>0} \frac{p(\mathbf{y} \circ x)}{(\mathbf{y} \circ x)^\alpha} = \frac{\mathbf{y}^\alpha}{p(\mathbf{y})} \cdot \text{Cap}_\alpha(p) = 1$$

which implies  $\nabla q(\mathbf{1}) = \alpha$ .

- $q$  is the **scaling** of  $p$  which makes the marginals equal to  $\alpha$ .
- **Now:**  $\text{Newt}(q) = \text{Newt}(p)$  and  $\nabla q(\mathbf{1}) = \alpha \implies \|\mathbf{1} - \alpha\|_1 \geq 2$ .

Works for any  $\alpha$  in relative interior of  $\text{Newt}(p)$ , so  $\|\mathbf{1} - \text{Newt}(p)\|_1 \geq 2$ .

**Scaling step:** Precomposing by  $\mathbf{y}$  updates “column sums”:

$$M \rightarrow \prod_{i=1}^n \sum_{j=1}^n m_{ij} x_j \rightarrow \prod_{i=1}^n \sum_{j=1}^n m_{ij} y_j x_j \rightarrow M \cdot \text{diag}(\mathbf{y}).$$

**Then:** Dividing by  $p(\mathbf{y})$  updates “row sums”:  $\prod_i \sum_j m_{ij} \cdot 1 = 1$ .

## An aside: Orbit closure of polynomial scaling

Fix real stable  $p(x_1, \dots, x_n)$  and scale by  $\mathbf{y} > 0$ :  $p \rightarrow q := \frac{p(\mathbf{y} \circ \mathbf{x})}{p(\mathbf{y})}$ .

**First:** Since  $\mathbf{y} > 0$ , we have that  $\text{supp}(q) = \text{supp}(p)$ .

**Question:** What supports can appear in the closure?

**Answer:** Precisely the collection of faces of  $\text{Newt}(p)$  intersected with  $\mathbb{Z}^n$ .

**Proof:** First, real stable polynomials have special “convexity property”:

- If  $\mathbf{x} \in \mathbb{Z}^n \cap \text{Newt}(p)$ , then  $\mathbf{x} \in \text{supp}(p)$ .

Fix  $p(\mathbf{x}) = \sum_{\mu} p_{\mu} \mathbf{x}^{\mu} \implies$  scaling by  $\mathbf{y}$  is  $q(\mathbf{x}) = \frac{\sum_{\mu} p_{\mu} \mathbf{y}^{\mu} \mathbf{x}^{\mu}}{\sum_{\mu} p_{\mu} \mathbf{y}^{\mu}}$ .

Consider  $\mathbf{y}_t := e^{t\mathbf{w}}$  for any  $\mathbf{w} \in \mathbb{R}^n \implies q_t(\mathbf{x}) = \frac{\sum_{\mu} p_{\mu} e^{t\langle \mathbf{w}, \mu \rangle} \mathbf{x}^{\mu}}{\sum_{\mu} p_{\mu} e^{t\langle \mathbf{w}, \mu \rangle}}$ .

As  $t \rightarrow \infty$ , terms which dominate are those for which  $\langle \mathbf{w}, \mu \rangle$  is maximized. Corresponds precisely to the faces of  $\text{Newt}(p)$  intersected with  $\text{supp}(p)$ .

**Similar situation for torus actions in general** (Levent’s talk yesterday).

- 1 Last time: Invariant-theoretic capacity
  - Null-cone problem
  - Invariant-theoretic capacity and the scaling-type algorithm
  - Scaling for real stable polynomials
- 2 Remembering all the forms of capacity
  - Polynomial capacity
  - Entropic capacity
  - Matrix capacity
  - Invariant-theoretic (non-commutative) capacity
- 3 Thanks!

# Polynomial capacity

Given a polynomial  $p \in \mathbb{R}_+[x_1, \dots, x_n]$  and  $\alpha \in \mathbb{R}_+^n$ , we define

$$\text{Cap}_\alpha(p) := \inf_{\mathbf{x} > 0} \frac{p(\mathbf{x})}{\mathbf{x}^\alpha} = \inf_{\mathbf{y} \in \mathbb{R}^n} \sum_{\mu \in \text{supp}(p)} p_\mu e^{\langle \mathbf{y}, \mu - \alpha \rangle}.$$

## Some basic properties:

- Log-convex program as function of  $\mathbf{y} \in \mathbb{R}^n$  (geometric programming).
- $\text{Cap}_\alpha(p) > 0$  iff  $\alpha \in \text{Newt}(p)$ .
  - **Kempf-Ness interpretation:**  $\alpha \in \text{Newt}(p)$  iff gradient =  $\mathbf{0}$  can be achieved when optimizing the objective.
  - **Entropy interpretation:** When  $p(\mathbf{1}) = 1$ ,  $-\log \text{Cap}_\alpha(p)$  is the entropy of a distribution on  $\text{supp}(p)$  with expectation  $\alpha$ .
- If  $p(\mathbf{1}) = 1$ , then  $\text{Cap}_\alpha(p) = 1$  iff  $\nabla p(\mathbf{1}) = \alpha$ .
  - **Entropy interpretation:**  $-\log \text{Cap}_\alpha(p)$  is the minimum relative entropy of a distribution on  $\text{supp}(p)$  with expectation  $\alpha$ .
  - **Scaling interpretation:** Cap was the measure of progress: once maximized, we are already scaled correctly.

# Capacity and real stable / Lorentzian polynomials

**Gurvits' theorem:** Real stable (or Lorentzian) polynomial  $p$  with non-negative coefficients in  $n$  variables of homogeneous degree  $n$ :

$$p_1 \geq \frac{n!}{n^n} \cdot \text{Cap}_1.$$

**Corollary:** Doubly stochastic  $p \implies p_1 \geq \frac{n!}{n^n}$  (permanent VdW bound).

**[Anari-Oveis Gharan], [Gurvits-L]:** Real stable  $p, q$  of degree at most  $\lambda_k$  in  $x_k$ , and  $\alpha \in \mathbb{R}_+^n$ :

$$\langle p, q \rangle = \sum_{\mu} \binom{\lambda}{\mu}^{-1} p_{\mu} q_{\mu} \geq K(\alpha, \lambda) \cdot \text{Cap}_{\alpha}(p) \text{Cap}_{\alpha}(q).$$

**[Anari-Liu-Oveis Gharan-Vinzant]:** Similar bound for Lorentzian / CLC.

**[Gurvits-L]:**  $T$  preserves real stability and  $\alpha, \beta \in \mathbb{R}_+^n$ :

$$\frac{\text{Cap}_{\beta}(T[p])}{\text{Cap}_{\alpha}(p)} \geq K(\alpha, \beta) \cdot \text{Cap}_{(\beta, \alpha)}(\text{Symb}[T]).$$

## Applications include bounds/approximations for:

- Permanent and mixed discriminant (Gurvits):  $T(p) = \partial_{x_n} |_{x_n=0} p$
- Contingency tables (Barvinok, Barvinok-Hartigan, Gurvits, Brändén-L-Pak) and Eulerian orientations (Csikvári-Schweitzer):

$$T(p) = \partial_{\mathbf{x}}^S p \quad \text{for specially chosen } S$$

- Biregular bipartite  $k$ -matchings (Gurvits-L):

$$T(p) = \sum_{S \in \binom{[n]}{k}} \partial_{\mathbf{x}}^S p$$

- Counting/optimization on stable matroids (Straszak-Vishnoi, Anari-Oveis Gharan) and intersection of two general matroids (Anari-Oveis Gharan-Vinzant):

$$T(p) = \sum_{B \in \mathcal{M}} \partial_{\mathbf{x}}^B p$$

# Entropic capacity

**Dual formulation** of polynomial capacity:  $p(\mathbf{1}) = 1$ , consider a distribution  $\mu$  on  $\text{supp}(p)$  given by  $\mu(\kappa) = p_\kappa$ . Then  $-\log \text{Cap}_\alpha(p) =$

$$\inf_{\mathbb{E}[\nu]=\alpha} D_{\text{KL}}(\nu\|\mu) = \inf_{\mathbb{E}[\nu]=\alpha} \sum_{\kappa \in \text{supp}(\mu)} \left[ \frac{\nu(\kappa)}{\mu(\kappa)} \log \frac{\nu(\kappa)}{\mu(\kappa)} \right] \mu(\kappa).$$

where  $\nu$  has support contained in  $\text{supp}(\mu)$ .

Allows for extension to **continuous case**: Fix a measure  $\mu$  supported on  $\Omega \subset \mathbb{R}^n$ , and pick some  $\theta \in \text{hull}(\Omega)$ :

$$\inf_{\substack{\nu=\phi \cdot \mu \\ \mathbb{E}[\nu]=\theta}} D_{\text{KL}}(\nu\|\mu) = \inf_{\mathbb{E}[\nu]=\theta} \int_{\Omega} \phi(\mathbf{x}) \log \phi(\mathbf{x}) d\mu(\mathbf{x}).$$

Define **continuous capacity** via:

$$\text{Cap}_\theta(\mu) := \inf_{\mathbf{x}>0} \frac{\int_{\Omega} \mathbf{x}^\kappa d\mu(\kappa)}{\mathbf{x}^\theta}.$$

As in the discrete case, this is dual to minimum relative entropy.



# Maximum entropy distributions

**Observation:** This shows that **scalings** give rise to maximum entropy (minimum relative entropy) distributions.

**Solution given by capacity:** If  $\alpha$  in the relative interior of  $\text{Newt}(p)$ , then there exists  $\mathbf{y} > 0$  (the scaling) such that  $\nu(\boldsymbol{\kappa}) = \frac{p_{\boldsymbol{\kappa}} \mathbf{y}^{\boldsymbol{\kappa}}}{p(\mathbf{y})}$ . That is,

$$q(\mathbf{x}) = \frac{p(\mathbf{y} \circ \mathbf{x})}{p(\mathbf{y})} = \sum_{\boldsymbol{\kappa}} \frac{p_{\boldsymbol{\kappa}} \mathbf{y}^{\boldsymbol{\kappa}}}{p(\mathbf{y})} \cdot \mathbf{x}^{\boldsymbol{\kappa}}$$

is the polynomial whose associated distribution is  $\nu$  (the KL optimum).

**What about continuous case?** If  $\theta$  in the relative interior of  $\text{hull}(\Omega)$ , then there exists  $\mathbf{y} > 0$  such that  $\phi(\boldsymbol{\kappa}) = \frac{\mathbf{y}^{\boldsymbol{\kappa}}}{\int_{\Omega} \mathbf{y}^{\boldsymbol{\gamma}} d\mu(\boldsymbol{\gamma})}$ . That is,

$$q(\mathbf{x}) = \frac{p(\mathbf{y} \circ \mathbf{x})}{p(\mathbf{y})} = \int_{\Omega} \frac{\mathbf{y}^{\boldsymbol{\kappa}}}{p(\mathbf{y})} \cdot \mathbf{x}^{\boldsymbol{\kappa}} d\mu(\boldsymbol{\kappa})$$

is the (torus) scaling of  $p(\mathbf{x}) = \int_{\Omega} \mathbf{x}^{\boldsymbol{\kappa}} d\mu(\boldsymbol{\kappa})$  with the desired marginals  $\theta$ .

# Matrix capacity

Used as a measure of progress for **operator scaling**:

$$\text{Cap}(T) := \inf_{X \succ 0} \frac{\det(T(X))}{\det(X)},$$

where  $T$  is completely positive or at least preserves PSD matrices.

**Analog to polynomial case:** Can we find positive definite  $Y$  such that:

$$\text{Cap}(T) = \frac{\det(T(Y))}{\det(Y)} \implies S(X) := \frac{T(Y^{1/2}XY^{1/2})}{T(Y)},$$

where  $S$  is the **scaling** of  $T$  to doubly stochastic. **Why?**

$$\text{Cap}(S) = \inf_{X \succ 0} \frac{\det(T(Y^{1/2}XY^{1/2}))}{\det(T(Y))\det(X)} = \inf_{X \succ 0} \frac{\det(T(X))}{\det(T(Y))\det(Y^{-1/2}XY^{-1/2})}$$

This equals  $\frac{\det(Y)}{\det(T(Y))} \cdot \text{Cap}(T) = 1 \iff S$  is doubly stochastic.

**See also [Franks]** for different marginals:  $\text{Cap}_A(p) = \inf_{X \succ 0} \frac{\det_A(T(X))}{\det_A(X)}$ .

# Invariant theoretic (non-commutative) capacity

**What we already saw at the beginning (and last week):** Let  $\pi : G \rightarrow \text{GL}(V)$  be a group representation, and define for  $v \in V$ :

$$\text{Cap}(v) = \text{Cap}_0(v) = \inf_{g \in G} \|\pi(g)v\|_2,$$

the minimum norm over the orbit of  $v$ .

In the **commutative (torus) case**, this is precisely a polynomial capacity problem. More generally in this case, **polynomial capacity**, **entropic capacity**, **invariant-theoretic capacity** all coincide. Also **maximum likelihood estimation** (see [Améndola-Kohn-Reichenbach-Seigal]), and **unconstrained geometric programming** (see [Bürgisser-Li-Nieuwboer-Walter]).

In the **left-right action case**, we know that **matrix capacity** and **invariant-theoretic capacity** coincide.

**Are there other connections between these notions of capacity?**

- 1 Last time: Invariant-theoretic capacity
  - Null-cone problem
  - Invariant-theoretic capacity and the scaling-type algorithm
  - Scaling for real stable polynomials
- 2 Remembering all the forms of capacity
  - Polynomial capacity
  - Entropic capacity
  - Matrix capacity
  - Invariant-theoretic (non-commutative) capacity
- 3 Thanks!

## **Thanks a bunch for joining on this course!**

Feel free to ask me about anything capacity-related or stability-related at any time. I am generally interested in any connections to these concepts.

Also feel free to give any feedback on the course. This is only the first of hopefully many times I will discuss this content. And I like to think of myself as someone who can take and utilize criticism, however harsh. So please feel free to test me on this!