Barriers for Tensor Scaling

Philipp Reichenbach

joint work with Cole Franks, [FR21], CCC 2021, arXiv:2102.06652

AGATES Workshop at IMPAN, Warsaw
Algebraic Geometry and Complexity Theory

16th November 2022

Outline

1. Introduction: setting, computational problems, motivation
    - Example: Matrix Scaling, Array Scaling
2. Scaling Algorithms and their Complexity
3. Bounds on Complexity Parameters
4. Glimpse of main Proof Techniques
5. Outlook and Open Problems

General Setting

- G is a reductive algebraic group over $\mathbb{C}$
  - e.g., $\mathrm{GL}(n)$, $(\mathbb{C}^\times)^n$, $\mathrm{SL}(n)$ and $\mathrm{ST}(n)$; products of these
- $\pi\colon G \to \mathrm{GL}(V)$ some finite-dim'l rational representation of G

General Setting

- G is a reductive algebraic group over $\mathbb{C}$
  - e.g., $\mathrm{GL}(n)$, $(\mathbb{C}^\times)^n$, $\mathrm{SL}(n)$ and $\mathrm{ST}(n)$; products of these
- $\pi\colon G \to \mathrm{GL}(V)$ some finite-dim'l rational representation of G

Concrete Example for this talk:

- *commutative:* $T = \mathrm{ST}(n)^d = \mathrm{ST}(n) \times \cdots \times \mathrm{ST}(n)$ (d copies)
- *non-commutative:* $G = \mathrm{SL}(n)^d$
- $V = (\mathbb{C}^n)^{\otimes d} = \mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n$ (think of $d = 2$ or $d = 3$)
- $\pi_{n,d}\colon G \to \mathrm{GL}(V)$ representation of G given by the action

$$(g_1, \ldots, g_d) \cdot v \mapsto (g_1 \otimes \cdots \otimes g_d)(v)$$

Computational Problems

- For fixed $v \in V$, consider $F_v \colon G \to \mathbb{R},\ g \mapsto \|g \cdot v\|^2$.
- The ***capacity*** of v is
  $\operatorname{cap}_G(v) := \inf_{g \in G} F_v(g) = \inf_{g \in G} \|g \cdot v\|^2$.

Computational Problems

- For fixed $v \in V$, consider $F_v \colon G \to \mathbb{R},\ g \mapsto \|g \cdot v\|^2$.
- The ***capacity*** of v is
  $\operatorname{cap}_G(v) := \inf_{g \in G} F_v(g) = \inf_{g \in G} \|g \cdot v\|^2$.
- **Norm Minimization:** Given $v \in V$ and $\varepsilon > 0$.
  Compute $g \in G$ such that $\|g \cdot v\|^2 \leq \operatorname{cap}_G(v) + \varepsilon$.
  - G commutative: *convex* optimization (geometric programming)
  - G non-commutative: *geodesic* convex optimization
  - **Desirable:** high precision efficiently, i.e., $\operatorname{poly}\big(n, \log(1/\varepsilon)\big)$

Computational Problems

- For fixed $v \in V$, consider $F_v \colon G \to \mathbb{R},\ g \mapsto \|g \cdot v\|^2$.
- The ***capacity*** of v is
  $\operatorname{cap}_G(v) := \inf_{g \in G} F_v(g) = \inf_{g \in G} \|g \cdot v\|^2$.
- **Norm Minimization:** Given $v \in V$ and $\varepsilon > 0$.
  Compute $g \in G$ such that $\|g \cdot v\|^2 \leq \operatorname{cap}_G(v) + \varepsilon$.
  - G commutative: *convex* optimization (geometric programming)
  - G non-commutative: *geodesic* convex optimization
  - **Desirable:** high precision efficiently, i.e., $\operatorname{poly}\big(n, \log(1/\varepsilon)\big)$
- **Null Cone Membership (NCM):**
  Given $v \in V$. *Decide* whether $\operatorname{cap}_G(v) = 0$.

Moment map and Kempf-Ness

- Fix $v \in V \backslash \{0\}$, consider $F_v \colon G \to \mathbb{R},\ g \mapsto \|g \cdot v\|^2$.
- **Moment map $\boldsymbol{\mu_G}$:**
  Think of $\mu_G(v)$ as the *gradient of* $\log F_v$ *at* $\mathrm{id} \in G$.

Moment map and Kempf-Ness

- Fix $v \in V \setminus \{0\}$, consider $F_v \colon G \to \mathbb{R},\ g \mapsto \|g \cdot v\|^2$.
- **Moment map $\boldsymbol{\mu_G}$:**
  Think of $\mu_G(v)$ as the *gradient of* $\log F_v$ *at* $\mathrm{id} \in G$.
- **Kempf-Ness:** $\mathrm{cap}_G(v) > 0 \quad \Leftrightarrow \quad \inf_{g \in G} \|\mu_G(g \cdot v)\| = 0$
  - (Geodesic) Convexity: minimum is attained iff gradient is zero
  - can be used to decide NCM

Moment map and Kempf-Ness

- Fix $v \in V \setminus \{0\}$, consider $F_v \colon G \to \mathbb{R},\ g \mapsto \|g \cdot v\|^2$.
- **Moment map $\boldsymbol{\mu_G}$:**
  Think of $\mu_G(v)$ as the *gradient of* $\log F_v$ *at* $\mathrm{id} \in G$.
- **Kempf-Ness:** $\mathrm{cap}_G(v) > 0 \quad \Leftrightarrow \quad \inf_{g \in G} \|\mu_G(g \cdot v)\| = 0$
  - (Geodesic) Convexity: minimum is attained iff gradient is zero
  - can be used to decide NCM
- **Scaling Problem:** Given $v \in V$ and $\varepsilon > 0$.
  Compute $g \in G$ such that $\|\mu_G(g \cdot v)\| \leq \varepsilon$.
- Quantitative version of Kempf-Ness relates Norm Minimzation and the Scaling Problem, [BFG+19]

Motivation: Versatile Applications

- **Convex Optimization:** Unconstrained geometric programming
- **Algebraic Geometry:** constructing moduli spaces
- **Analysis:** Brascamp-Lieb inequalities
- **Physics:** quantum information theory
- **Computer Science:** non-commutative RIT
- **Geometric Complexity Theory**
- **(Algebraic) Statistics:** matrix scaling; ML estimation for Gaussian group models and for log-linear models [AKRS21]

Motivation: Versatile Applications

- $\mathrm{T} = \mathrm{ST}(n)^d$ commutative; $G = \mathrm{SL}(n)^d$ non-commutative
- natural action on $(\mathbb{C}^n)^{\otimes d}$ via $g_1 \otimes \cdots \otimes g_d$
- $(\mathbb{C}^n)^{\otimes d}$: think of matrices ($d = 2$) or 3-tensors ($d = 3$)

Motivation: Versatile Applications

- $\mathrm{T} = \mathrm{ST}(n)^d$ commutative; $G = \mathrm{SL}(n)^d$ non-commutative
- natural action on $(\mathbb{C}^n)^{\otimes d}$ via $g_1 \otimes \cdots \otimes g_d$
- $(\mathbb{C}^n)^{\otimes d}$: think of matrices ($d = 2$) or 3-tensors ($d = 3$)

	T: commutative	G: non-commutative
$d = 2$	**matrix scaling:** statistics, bipartite matching, optimal transport, ...	**operator scaling:** non-commutative RIT, Tyler's M estimator, matrix normal models, ...
$d = 3$	**array scaling:** multimarginal transport, ...	**tensor scaling:** geometric complexity theory, quantum information theory, tensor normal models, ...

Digression: Matrix Scaling

- A non-negative matrix $A \in \mathbb{R}_{\geq 0}^{n \times n}$ is called **doubly stochastic**, if each of its row sums and each of its column sums equals one.
- For *positive diagonal* matrices X and Y, we say XAY is a **scaling** of A.
- **Matrix Scaling:** Given non-negative matrix A.
  Can we scale A arbitrarily close to doubly stochastic?

$$XAY = \begin{pmatrix} e^{x_1} & & \\ & \ddots & \\ & & e^{x_n} \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} e^{y_1} & & \\ & \ddots & \\ & & e^{y_n} \end{pmatrix}$$

where $x, y \in \mathbb{R}^n$.

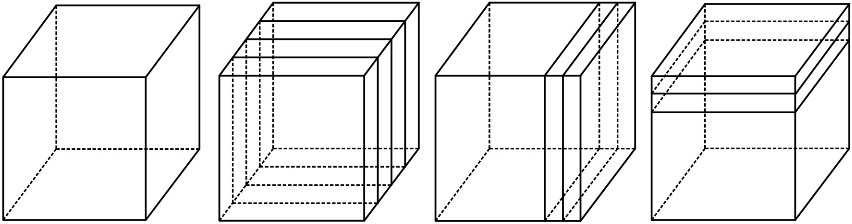
Digression: Matrix Scaling

- A non-negative matrix $A \in \mathbb{R}^{n \times n}_{\geq 0}$ is called **doubly stochastic**, if each of its row sums and each of its column sums equals one.
- For *positive diagonal* matrices X and Y, we say XAY is a **scaling** of A.
- **Matrix Scaling:** Given non-negative matrix A.
  Can we scale A arbitrarily close to doubly stochastic?

$$\begin{pmatrix} \varepsilon & & \\ & \varepsilon^{-1} & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon & & \\ & \varepsilon^{-1} & \\ & & 1 \end{pmatrix} = \begin{pmatrix} \varepsilon^2 & 1 & \varepsilon \\ 1 & 0 & 0 \\ \varepsilon & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Glimpse of Array Scaling

- Given an array $A \in (\mathbb{R}^n_{\geq 0})^{\otimes 3}$. Scaling of A is $(X \otimes Y \otimes Z)(A)$, where X, Y and Z are positive diagonal matrices.
- **Array Scaling:** Can we scale A in the limit to tristochastic?
- **tristochastic:** all slice sums are equal to one



Matrix Scaling as a geometric program

- For $\mathrm{T} = \mathrm{ST}(n)^2$ and a matrix $v \in \mathbb{C}^{n \times n}$ we get

$$\mathrm{cap}_{\mathrm{T}}(v) = \cdots = \inf_{x,y \in \mathbb{R}^n} f_v(x,y) := \inf_{x,y \in \mathbb{R}^n} \sum_{i,j=1}^{n} |v_{ij}|^2 \, e^{(\varepsilon_i, \varepsilon_j) \cdot (x,y)}$$

where $\varepsilon_i := e_i - \frac{1}{n} \mathbb{1}_n$. This is matrix scaling for $A = (|v_{ij}|^2)_{i,j}$!

Matrix Scaling as a geometric program

- For $\mathrm{T} = \mathrm{ST}(n)^2$ and a matrix $v \in \mathbb{C}^{n \times n}$ we get

$$\operatorname{cap}_{\mathrm{T}}(v) = \cdots = \inf_{x,y \in \mathbb{R}^n} f_v(x,y) := \inf_{x,y \in \mathbb{R}^n} \sum_{i,j=1}^{n} |v_{ij}|^2 \, e^{(\varepsilon_i, \varepsilon_j) \cdot (x,y)}$$

  where $\varepsilon_i := e_i - \frac{1}{n} \mathbb{1}_n$. This is matrix scaling for $A = (|v_{ij}|^2)_{i,j}$!

- **Fact:** A is approximately scalable if and only if $\operatorname{cap}_{\mathrm{T}}(v) > 0$.
  - **NCM:** Decide whether A is *not* approximately scalable.

Matrix Scaling as a geometric program

- For $\mathrm{T} = \mathrm{ST}(n)^2$ and a matrix $v \in \mathbb{C}^{n \times n}$ we get

$$\mathrm{cap}_{\mathrm{T}}(v) = \cdots = \inf_{x,y \in \mathbb{R}^n} f_v(x,y) := \inf_{x,y \in \mathbb{R}^n} \sum_{i,j=1}^{n} |v_{ij}|^2 \, e^{(\varepsilon_i, \varepsilon_j) \cdot (x,y)}$$

  where $\varepsilon_i := e_i - \frac{1}{n} \mathbb{1}_n$. This is matrix scaling for $A = (|v_{ij}|^2)_{i,j}$!

- **Fact:** A is approximately scalable if and only if $\mathrm{cap}_{\mathrm{T}}(v) > 0$.
  - **NCM:** Decide whether A is *not* approximately scalable.

- $\|\mu_{\mathrm{T}}(g \cdot v)\| = \|\nabla \log \big(f_v(x,y)\big)\|$ measures the deviation of XAY from being doubly stochastic.

- **Scaling Problem:** (closely related to Norm Minimization)
  Compute scaling of A that is ε-close to doubly stochastic.

Matrix Scaling: important concepts

- The **weight polytope** of v is $\Delta_{\mathrm{T}}(v) := \operatorname{conv}\{(\varepsilon_i, \varepsilon_j) \mid v_{ij} \neq 0\}$.
- **Fact:** $\operatorname{cap}_{\mathrm{T}}(v) > 0 \quad \Leftrightarrow \quad 0 \in \Delta_{\mathrm{T}}(v)$.
- The function $(x, y) \mapsto \log\big(f_v(x, y)\big)$ is convex.
- $\|\mu_{\mathrm{T}}(g \cdot v)\| = \|\nabla \log\big(f_v(x, y)\big)\|$ measures the deviation of XAY from being doubly stochastic.

Matrix Scaling: important concepts

- The **weight polytope** of v is $\Delta_{\mathrm{T}}(v) := \operatorname{conv}\{(\varepsilon_i, \varepsilon_j) \mid v_{ij} \neq 0\}$.
- **Fact:** $\operatorname{cap}_{\mathrm{T}}(v) > 0 \quad \Leftrightarrow \quad 0 \in \Delta_{\mathrm{T}}(v)$.
- The function $(x, y) \mapsto \log\big(f_v(x, y)\big)$ is convex.
- $\|\mu_{\mathrm{T}}(g \cdot v)\| = \|\nabla \log\big(f_v(x, y)\big)\|$ measures the deviation of XAY from being doubly stochastic.
- Concepts generalize to $d = 3$ and to G non-commutative
- For $d = 3$, array scaling arises for a tensor $v = (v_{ijk})_{i,j,k}$ as

$$\operatorname{cap}_{\mathrm{T}}(v) = \inf_{x,y,z \in \mathbb{R}^n} f_v(x, y, z) := \inf_{x,y,z \in \mathbb{R}^n} \sum_{i,j,k=1}^{n} |v_{ijk}|^2 \, e^{(\varepsilon_i, \varepsilon_j, \varepsilon_k) \cdot (x,y,z)}$$

Questions?

Next: 2. Scaling Algorithms and their Complexity

Scaling Algorithms and their Complexity

HP: efficient computation of high precision solutions, i.e., in $\mathrm{poly}\left(n, \log(1/\varepsilon)\right)$ time, for norm minimization problem

NCM: null cone membership solvable in poly-time

	T: commutative	G: non-commutative
$d = 2$	**matrix scaling:** HP, NCM (trust region, IPM)	**operator scaling:** HP, NCM (trust region)
$d = 3$	**array scaling:** HP, NCM (IPM, *not* via trust region)	**tensor scaling:** HP, NCM (no IPM available)

IPM is a shorthand for interior point methods

Natural Questions

Can we explain the dichotomy between $d = 2$ and $d = 3$?

- Why do gradient descent and trust region methods not suffice for HP and NCM when $d = 3$?
- Are known algorithms good enough for tensor scaling and only the complexity analysis lacks to show this?
  Or do we need new algorithmic approaches?

Natural Questions

Can we explain the dichotomy between $d = 2$ and $d = 3$?

- Why do gradient descent and trust region methods not suffice for HP and NCM when $d = 3$?
- Are known algorithms good enough for tensor scaling and only the complexity analysis lacks to show this?
  Or do we need new algorithmic approaches?

To answer these questions we investigate

- for HP: *diameter* of approximate minimizers
- for NCM: *precision parameters* (*weight margin*, *gap*)

Questions?

Next: 3. Bounds on Complexity Parameters

Part I: Diameter

Diameter

Norm minimization: Given a tensor $v = (v_{ijk})_{i,j,k}$ and $\varepsilon > 0$.
Find $g \in G$ such that $\|g \cdot v\|^2 \leq \mathrm{cap}(v) + \varepsilon$.

Think of the **diameter** $D_v(\varepsilon)$ as

- (bit) complexity of approximate minimizer
- smallest distance in G between id and approximate minimizer g

Diameter

Norm minimization: Given a tensor $v = (v_{ijk})_{i,j,k}$ and $\varepsilon > 0$.
Find $g \in G$ such that $\|g \cdot v\|^2 \leq \mathrm{cap}(v) + \varepsilon$.

Think of the **diameter** $D_v(\varepsilon)$ as

- (bit) complexity of approximate minimizer
- smallest distance in G between id and approximate minimizer g

In the commutative case use

$$\mathrm{cap}(v) = \inf_{x,y,z \in \mathbb{R}^n} f_v(x,y,z) := \inf_{x,y,z \in \mathbb{R}^n} \sum_{i,j,k=1}^{n} |v_{ijk}|^2 \, e^{(\varepsilon_i, \varepsilon_j, \varepsilon_k) \cdot (x,y,z)}$$

and define $\boldsymbol{D_v(\varepsilon)}$ as the infimum over all $R > 0$ such that

$$\inf_{\|(x,y,z)\| \leq R} f_v(x,y,z) \leq \mathrm{cap}(v) + \varepsilon.$$

Diameter

- $D_v(\varepsilon)$ can be defined in an analogous way in the non-commutative setting

Diameter

- $D_v(\varepsilon)$ can be defined in an analogous way in the non-commutative setting

Known upper bounds on $D_v(\varepsilon)$:

	T: commutative	G: non-commutative
$d = 2$	**matrix scaling:** $O(n \log(1/\varepsilon))$	**operator scaling:** $O\big(n^{3/2} \log(1/\varepsilon)\big)$
$d = 3$	**array scaling:** [SV19] $O\big(n^{3/2} 2^{6n} \log(1/\varepsilon)\big)$	**tensor scaling:** [BFG+19] $O\big(2^{n \log(n)} \log(1/\varepsilon)\big)$

Diameter: Main Theorem

Theorem: Diameter bound for tensor scaling [Franks, R.]

There exists a constant $C > 0$ such that for all $\varepsilon \leq \exp(-Cn^2 \log n)$, there is a tensor $v = v(\varepsilon)$ with

$$D_v(\varepsilon) = \Omega\left(2^{n/3} \log(1/\varepsilon)\right).$$

Diameter: Main Theorem

Theorem: Diameter bound for tensor scaling [Franks, R.]

There exists a constant $C > 0$ such that for all $\varepsilon \leq \exp(-Cn^2 \log n)$, there is a tensor $v = v(\varepsilon)$ with

$$D_v(\varepsilon) = \Omega\left(2^{n/3} \log(1/\varepsilon)\right).$$

Here, v is **reasonably nice**:
v has $O(n)$ nonzero entries of bit complexity $O(\log n + \log(1/\varepsilon))$, $1/4 \leq \operatorname{cap}(v) \leq 1$ and $1/2 \leq \|v\| \leq 1$.

A similar theorem holds for array scaling (commutative and $d = 3$).

$\rightsquigarrow$ There we can drop the dependence of v on ε.

Diameter: Implications of Main Theorem

- algorithms with constant (or poly-bounded!) step size cannot provide HP for array/tensor scaling, e.g.,
  - gradient descent
  - trust region methods
- **commutative** (array scaling): our construction answers an open problem of [SV19] in the affirmative
- **non-commutative** (tensor scaling): together with a result of [HM21] our construction suggests that even cutting plane methods (as described by [Rus20]) do not suffice for HP

Questions?

Next: 3. Bounds on Complexity Parameters

Part II: Precision Parameters

Precision Parameters

- **NCM:** For $v \in V$, decide whether $\mathrm{cap}(v) = 0$.
- "global" parameters $\gamma_{\mathrm{T}}(\pi)$ (**weight margin**) and $\gamma_G(\pi)$ (**gap**)
  - capture required precision to solve NCM for T resp. G
  - always have $\gamma_{\mathrm{T}}(\pi) \leq \gamma_G(\pi)$

Precision Parameters

- **NCM:** For $v \in V$, decide whether $\operatorname{cap}(v) = 0$.
- "global" parameters $\gamma_{\mathrm{T}}(\pi)$ (**weight margin**) and $\gamma_G(\pi)$ (**gap**)
  - capture required precision to solve NCM for T resp. G
  - always have $\gamma_{\mathrm{T}}(\pi) \leq \gamma_G(\pi)$
- $\gamma_{\mathrm{T}}(\pi)^{-1}$ appears polynomially in running time bounds, [BFG+19]
- Diameter bound: $D_v(\varepsilon) \leq \operatorname{poly}\left(\gamma_{\mathrm{T}}(\pi)^{-1}, \log(1/\varepsilon)\right)$, [BFG+19]
- **Quantitative Kempf-Ness [BFG+19]:** For $v \neq 0$

$$1 - \frac{\|\mu_G(v)\|_F}{\gamma_{\mathrm{T}}(\pi)} \leq \frac{\operatorname{cap}_G(v)}{\|v\|^2} \left(\leq 1 - \frac{\|\mu_G(v)\|^2}{4\,\mathrm{N}(\pi)^2} \right).$$

Weight margin: Formal Definition

- **Precision Parameter:**
  - $\gamma_{\mathrm{T}}(\pi) := \min \left\{ \|\mu_{\mathrm{T}}(v)\| \mid v \neq 0,\ \mathrm{cap}_{\mathrm{T}}(v) = 0 \right\}$
  - $\gamma_{\mathrm{T}}(\pi)$ is the largest constant with the property: Whenever $\|\mu_{\mathrm{T}}(t \cdot v)\| < \gamma_{\mathrm{T}}(\pi)$, then $\mathrm{cap}_{\mathrm{T}}(v) > 0$.

Weight margin: Formal Definition

- **Precision Parameter:**
  - $\gamma_{\mathrm{T}}(\pi) := \min \left\{ \|\mu_{\mathrm{T}}(v)\| \mid v \neq 0,\ \mathrm{cap}_{\mathrm{T}}(v) = 0 \right\}$
  - $\gamma_{\mathrm{T}}(\pi)$ is the largest constant with the property:
    Whenever $\|\mu_{\mathrm{T}}(t \cdot v)\| < \gamma_{\mathrm{T}}(\pi)$, then $\mathrm{cap}_{\mathrm{T}}(v) > 0$.
- **Combinatorial Description:**
  Matrix scaling: set of weights is $\Omega = \{(\varepsilon_i, \varepsilon_j) \mid i, j \in [n]\} \subseteq \mathbb{R}^{2n}$

$$\begin{aligned} \gamma_{\mathrm{T}}(\pi) &= \min \left\{ \operatorname{dist}\big(0, \operatorname{conv}(S)\big) \mid S \subseteq \Omega, 0 \notin \operatorname{conv}(S) \right\} \\ &= \min \left\{ \operatorname{dist}\big(0, \Delta_{\mathrm{T}}(v)\big) \mid v \neq 0,\ 0 \notin \Delta_{\mathrm{T}}(v) \right\} \end{aligned}$$

Weight margin: Formal Definition

- **Precision Parameter:**
  - $\gamma_{\mathrm{T}}(\pi) := \min\left\{ \|\mu_{\mathrm{T}}(v)\| \mid v \neq 0,\ \mathrm{cap}_{\mathrm{T}}(v) = 0 \right\}$
  - $\gamma_{\mathrm{T}}(\pi)$ is the largest constant with the property: Whenever $\|\mu_{\mathrm{T}}(t \cdot v)\| < \gamma_{\mathrm{T}}(\pi)$, then $\mathrm{cap}_{\mathrm{T}}(v) > 0$.
- **Combinatorial Description:**
  Matrix scaling: set of weights is $\Omega = \{(\varepsilon_i, \varepsilon_j) \mid i, j \in [n]\} \subseteq \mathbb{R}^{2n}$

$$\begin{aligned} \gamma_{\mathrm{T}}(\pi) &= \min\left\{ \mathrm{dist}\left(0, \mathrm{conv}(S)\right) \mid S \subseteq \Omega, 0 \notin \mathrm{conv}(S) \right\} \\ &= \min\left\{ \mathrm{dist}\left(0, \Delta_{\mathrm{T}}(v)\right) \mid v \neq 0,\ 0 \notin \Delta_{\mathrm{T}}(v) \right\} \end{aligned}$$

- Analogously for $\gamma_G(\pi)$ via μ_G and moment polytopes $\Delta_G(v)$

Weight margin and Gap: Interesting Example

Let $\tau_{n,d}$ be the natural action of $G = \mathrm{SL}(n)^d$ on the quiver Q_d with dimension vector $(n, \ldots, n)$, where Q_d is

$$1 \longleftarrow 2 \longrightarrow 3 \cdots\cdots\cdots\; d-2 \longrightarrow d-1 \longleftarrow d.$$

Weight margin and Gap: Interesting Example

Let $\tau_{n,d}$ be the natural action of $G = \mathrm{SL}(n)^d$ on the quiver Q_d with dimension vector $(n, \ldots, n)$, where Q_d is

$$1 \longleftarrow 2 \longrightarrow 3 \cdots\cdots\cdots d-2 \longrightarrow d-1 \longleftarrow d.$$

Theorem: small Weight margin and large Gap

For $n, d \geq 2$ it holds that

(a) $\gamma_{\mathrm{T}}(\tau_{n,d}) \leq (n-1)^{-d+1}$ [Franks, R.]

(b) $\gamma_G(\tau_{n,d}) \gtrsim (dn)^{-1}$ [Franks, Makam]

Weight margin and Gap: Interesting Example

Let $\tau_{n,d}$ be the natural action of $G = \mathrm{SL}(n)^d$ on the quiver Q_d with dimension vector $(n, \ldots, n)$, where Q_d is

$$1 \longleftarrow 2 \longrightarrow 3 \cdots\cdots\cdots d-2 \longrightarrow d-1 \longleftarrow d.$$

Theorem: small Weight margin and large Gap

For $n, d \geq 2$ it holds that

(a) $\gamma_{\mathrm{T}}(\tau_{n,d}) \leq (n-1)^{-d+1}$ [Franks, R.]

(b) $\gamma_G(\tau_{n,d}) \gtrsim (dn)^{-1}$ [Franks, Makam]

(c) $\gamma_G\left(\tau_{n,d}^{\oplus n}\right) \leq (n-1)^{-d+1}$ [Franks, R.]

Note: $\tau_{n,d}^{\oplus k}$ corresponds to the action of G on Q_d, but each arrow appears k-many times.

Weight margin and Gap: Main Theorem

Known lower bounds:

- for $d = 2$: $\Omega\big(n^{-3/2}\big) = \gamma_{\mathrm{T}}(\pi_{n,2}) \le \gamma_G(\pi_{n,2})$ [LSW98; Gur04]
- for $d = 3$: $\Omega\big((\sqrt{3}n)^{-3n-1}\big) = \gamma_{\mathrm{T}}(\pi_{n,3}) \le \gamma_G(\pi_{n,3})$ [BFG+19]

Weight margin and Gap: Main Theorem

Known lower bounds:

- for $d = 2$: $\Omega\left(n^{-3/2}\right) = \gamma_{\mathrm{T}}(\pi_{n,2}) \leq \gamma_G(\pi_{n,2})$ [LSW98; Gur04]
- for $d = 3$: $\Omega\left((\sqrt{3}n)^{-3n-1}\right) = \gamma_{\mathrm{T}}(\pi_{n,3}) \leq \gamma_G(\pi_{n,3})$ [BFG+19]

Theorem: Bound for array/tensor scaling [Franks, R.]

There is a constant $C > 0$ such that for all $n \geq 2$, $d \geq 3$

$$\gamma_{\mathrm{T}}(\pi_{n,d}) \leq \gamma_G(\pi_{n,d}) \leq 2^{-Cnd}.$$

Weight margin and Gap: Main Theorem

Known lower bounds:

- for $d = 2$: $\Omega\big(n^{-3/2}\big) = \gamma_{\mathrm{T}}(\pi_{n,2}) \leq \gamma_G(\pi_{n,2})$ [LSW98; Gur04]
- for $d = 3$: $\Omega\big((\sqrt{3}n)^{-3n-1}\big) = \gamma_{\mathrm{T}}(\pi_{n,3}) \leq \gamma_G(\pi_{n,3})$ [BFG+19]

Theorem: Bound for array/tensor scaling [Franks, R.]

There is a constant $C > 0$ such that for all $n \geq 2$, $d \geq 3$

$$\gamma_{\mathrm{T}}(\pi_{n,d}) \leq \gamma_G(\pi_{n,d}) \leq 2^{-Cnd}.$$

Bound explains:

- why current methods cannot give NCM for tensor scaling
- why diameter can be exponentially large for array/tensor scaling

Weight margin and Gap: Detailed Bounds

Theorem: Detailed bounds [Franks, R.]

(a) For $n = 2$ and $d \geq 3$, $\gamma_{\mathrm{T}}(\pi_{2,d}) \leq \gamma_G(\pi_{2,d}) \leq 2^{-\frac{d}{2}+1} = 2^{-\frac{nd}{4}+1}$.

- [AV97]: $\gamma_{\mathrm{T}}(\pi_{2,d}) = 2^{-\Theta(d \log d)}$
  - not easy to check whether their construction is *free*
- [MS18]: Qubits $v \in (\mathbb{C}^2)^{\otimes d}$ witnessing the exponentially small behaviour of $\gamma_G(\pi_{2,d})$ are rare

Weight margin and Gap: Detailed Bounds

Theorem: Detailed bounds [Franks, R.]

(a) For $n = 2$ and $d \geq 3$, $\gamma_{\mathrm{T}}(\pi_{2,d}) \leq \gamma_G(\pi_{2,d}) \leq 2^{-\frac{d}{2}+1} = 2^{-\frac{nd}{4}+1}$.

(b) For $n \geq 3$ and $d = 3$, $\gamma_{\mathrm{T}}(\pi_{n,3}) \leq \gamma_G(\pi_{n,3}) \leq 2^{-n+1} = 2^{-\frac{nd}{3}+1}$.

(c) If $n \geq 3$ and $d = 6r - 3$ for some integer $r \geq 2$, then

$$\gamma_{\mathrm{T}}(\pi_{n,d}) \leq \gamma_G(\pi_{n,d}) \leq 2^{-r(n-1)+1} \approx 2^{-\frac{nd}{6}}.$$

Weight margin and Gap: Detailed Bounds

Theorem: Detailed bounds [Franks, R.]

(a) For $n = 2$ and $d \geq 3$, $\gamma_{\mathrm{T}}(\pi_{2,d}) \leq \gamma_G(\pi_{2,d}) \leq 2^{-\frac{d}{2}+1} = 2^{-\frac{nd}{4}+1}$.

(b) For $n \geq 3$ and $d = 3$, $\gamma_{\mathrm{T}}(\pi_{n,3}) \leq \gamma_G(\pi_{n,3}) \leq 2^{-n+1} = 2^{-\frac{nd}{3}+1}$.

(c) If $n \geq 3$ and $d = 6r - 3$ for some integer $r \geq 2$, then

$$\gamma_{\mathrm{T}}(\pi_{n,d}) \leq \gamma_G(\pi_{n,d}) \leq 2^{-r(n-1)+1} \approx 2^{-\frac{nd}{6}}.$$

- Above results can be padded to obtain previous theorem.
  - constant $C > 0$ is reasonable (think of $C \approx \frac{1}{6}$ for $n, d \gg 0$)
- can be used to obtain exponentially small upper bounds on $\gamma_{\mathrm{T}}(\pi)$ and $\gamma_G(\pi)$ for *polynomial scaling* [Franks, R.]

Questions?

Next: 4. Glimpse of main Proof Techniques

Proof Ideas: commutative case

- Both, diameter bounds and weight margin bounds are proven by exploiting **extremal combinatorics** (of set of weights $\Omega(\pi_{n,d})$)
- **Intuition for $\boldsymbol{\gamma_{\mathrm{T}}(\pi)}$:** Birkhoff polytope B_2 of doubly stochastic matrices behaves well, while its 3-dim'l analogue B_3 does not.

$$B_2 = \left\{ (x_{ij}) \in \left(\mathbb{R}^{n \times n}_{\geq 0} \right) \mid \forall i_0, j_0 \in [n] \colon \sum_j x_{i_0,j} = 1, \sum_i x_{i,j_0} = 1 \right\}$$

$$B_3 = \left\{ (x_{ijk}) \in \left(\mathbb{R}^{n}_{\geq 0} \right)^{\otimes 3} \mid \forall i_0 \in [n] \colon \sum_{j,k} x_{i_0,j,k} = 1 \text{ etc.} \right\}$$

Proof Ideas: commutative case

- Both, diameter bounds and weight margin bounds are proven by exploiting **extremal combinatorics** (of set of weights $\Omega(\pi_{n,d})$)
- **Intuition for $\boldsymbol{\gamma_{\mathbf{T}}(\pi)}$:** Birkhoff polytope B_2 of doubly stochastic matrices behaves well, while its 3-dim'l analogue B_3 does not.

$$B_2 = \left\{ (x_{ij}) \in \left(\mathbb{R}^{n \times n}_{\geq 0}\right) \mid \forall i_0, j_0 \in [n] \colon \sum_j x_{i_0,j} = 1, \sum_i x_{i,j_0} = 1 \right\}$$

$$B_3 = \left\{ (x_{ijk}) \in \left(\mathbb{R}^{n}_{\geq 0}\right)^{\otimes 3} \mid \forall i_0 \in [n] \colon \sum_{j,k} x_{i_0,j,k} = 1 \text{ etc.} \right\}$$

- **E.g., for $\boldsymbol{\gamma_{\mathbf{T}}(\pi_{n,3})}$:** we crucially use a result of [Kra07], that explicitly states a vertex of B_3 with an exponentially small entry $\rightsquigarrow$ gives $\Gamma_{n,3} \subseteq \Omega(\pi_{n,3})$ with $0 < \operatorname{dist}\big(0, \operatorname{conv}(\Gamma_{n,3})\big) \leq 2^{-n+1}$

Proof Ideas: non-commutative case

- **Problem:** We act by general (and not just diagonal) matrices.
- **Solution:** Notion of a *free* tensor

Proof Ideas: non-commutative case

- **Problem:** We act by general (and not just diagonal) matrices.
- **Solution:** Notion of a *free* tensor

Definition: *free* tensor $v = (v_{ijk})_{i,j,k}$

Whenever $v_{ijk} \neq 0$ and $v_{i'j'k'} \neq 0$, then the triples (i, j, k) and (i', j', k') are *either* the same *or* differ in at least two entries.

Proof Ideas: non-commutative case

- **Problem:** We act by general (and not just diagonal) matrices.
- **Solution:** Notion of a *free* tensor

Definition: *free* tensor $v = (v_{ijk})_{i,j,k}$

Whenever $v_{ijk} \neq 0$ and $v_{i'j'k'} \neq 0$, then the triples (i,j,k) and (i',j',k') are *either* the same *or* differ in at least two entries.

- appears as *strong orthogonal* in [DK85]
- used to study moment maps and/or moment polytopes, e.g., in [Sja98], [Fra02], [MS15] and [CVZ18]

Freeness for 3-tensors

Definition: *free* tensor $v = (v_{ijk})_{i,j,k}$

Whenever $v_{ijk} \neq 0$ and $v_{i'j'k'} \neq 0$, then the triples (i,j,k) and (i',j',k') are *either* the same *or* differ in at least two entries.

- Given $v \in (\mathbb{C}^n)^{\otimes 3} \setminus \{0\}$, let $M^{(1)}, M^{(2)}, M^{(3)} \in \mathbb{C}^{n \times n^2}$ be the *flattenings* of v, e.g., $M^{(1)}_{i,(j,k)} = v_{ijk}$. Then

$$\mu_G(v) = \left(M^{(l)} \left(M^{(l)} \right)^\dagger - \frac{\|M^{(l)}\|_F^2}{n} I_n \right)_{l=1,2,3}$$

- If v is free, then $\mu_G(v)$ diagonal. E.g., for $s,t \in [n]$, $s \neq t$

$$\left(M^{(1)} \left(M^{(1)} \right)^\dagger \right)_{s,t} = \sum_{j,k=1}^{n} M^{(1)}_{s,(j,k)} \overline{M^{(1)}_{t,(j,k)}} = \sum_{j,k=1}^{n} v_{s,j,k} \overline{v_{t,j,k}} \overset{\text{free}}{=} 0.$$

Freeness for 3-tensors

Definition: *free* tensor $v = (v_{ijk})_{i,j,k}$

Whenever $v_{ijk} \neq 0$ and $v_{i'j'k'} \neq 0$, then the triples (i,j,k) and (i',j',k') are *either* the same *or* differ in at least two entries.

- Freeness ensures that *"all information is contained in the action with diagonal matrices"*
  - can lift free constructions from the commutative to the non-commutative case
- All our constructions are free.

Questions?

Next: 5. Outlook and Open Problems

Outlook and Open Problems

- Is $\varepsilon \leq \exp(-Cn^2 \log n)$ necessary for our diameter bound? Or are there similar diameter bounds for, e.g., $\varepsilon \leq \exp(-Cn)$?
- Are the upper bounds on $\gamma_{\mathrm{T}}(\pi_{n,d})$ and $\gamma_G(\pi_{n,d})$ essentially tight?
- Large gap $\gamma_G(\pi)$ for *all* quivers of finite type?
- Quantitative Kempf-Ness via the gap $\gamma_G(\pi)$? (instead of $\gamma_{\mathrm{T}}(\pi)$)

Outlook and Open Problems

- Is $\varepsilon \leq \exp(-Cn^2 \log n)$ necessary for our diameter bound? Or are there similar diameter bounds for, e.g., $\varepsilon \leq \exp(-Cn)$?
- Are the upper bounds on $\gamma_{\mathrm{T}}(\pi_{n,d})$ and $\gamma_G(\pi_{n,d})$ essentially tight?
- Large gap $\gamma_G(\pi)$ for *all* quivers of finite type?
- Quantitative Kempf-Ness via the gap $\gamma_G(\pi)$? (instead of $\gamma_{\mathrm{T}}(\pi)$)
- **Upshot:** Our results show that for *tensor scaling* current methods *neither* suffice for HP *nor* for NCM.

  Therefore, our work strongly motivates:
- **Major Goal:** Find (interior point) methods for geodesic convex setting with running time $\mathrm{poly}\big(\boldsymbol{\log(1/\gamma(\pi))}, \log(1/\varepsilon)\big)$.

Thank you!

References I

[AKRS21] C. Améndola, K. Kohn, P. Reichenbach, and A. Seigal. "Invariant Theory and Scaling Algorithms for Maximum Likelihood Estimation". In: *SIAM J. Appl. Algebra Geom.* 5.2 (2021), pp. 304–337. doi: 10.1137/20M1328932.

[AV97] N. Alon and V. H. Vũ. "Anti-Hadamard matrices, coin weighing, threshold gates, and indecomposable hypergraphs". In: *J. Combin. Theory Ser. A* 79.1 (1997), pp. 133–160. doi: 10.1006/jcta.1997.2780.

[BFG+19] P. Bürgisser, C. Franks, A. Garg, R. Oliveira, M. Walter, and A. Wigderson. *Towards a theory of non-commutative optimization: geodesic 1st and 2nd order methods for moment maps and polytopes*. 2019. doi: 10.48550/ARXIV.1910.12375.

[CVZ18] M. Christandl, P. Vrana, and J. Zuiddam. "Universal points in the asymptotic spectrum of tensors". In: *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. ACM, New York, 2018, pp. 289–296. doi: 10.1145/3188745.3188766.

References II

[DK85] J. Dadok and V. Kac. “Polar representations”. In: *J. Algebra* 92.2 (1985), pp. 504–524. doi: 10.1016/0021-8693(85)90136-X.

[FR21] W. C. Franks and P. Reichenbach. “Barriers for recent methods in geodesic optimization”. In: *36th Computational Complexity Conference*. Vol. 200. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2021, Art. No. 13, 54. doi: 10.4230/LIPIcs.CCC.2021.13.

[Fra02] M. Franz. “Moment polytopes of projective G-varieties and tensor products of symmetric group representations”. In: *J. Lie Theory* 12.2 (2002), pp. 539–549.

[Gur04] L. Gurvits. “Classical complexity and quantum entanglement”. In: *J. Comput. System Sci.* 69.3 (2004), pp. 448–484. doi: 10.1016/j.jcss.2004.06.003.

[HM21] L. Hamilton and A. Moitra. “No-go Theorem for Acceleration in the Hyperbolic Plane”. 2021.

References III

[Kra07] V. M. Kravtsov. “Combinatorial properties of noninteger vertices of a polytope in a three-index axial assignment problem”. In: *Kibernet. Sistem. Anal.* 43.1 (2007), pp. 33–44, 189. doi: 10.1007/s10559-007-0023-0.

[LSW98] N. Linial, A. Samorodnitsky, and A. Wigderson. “A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents”. In: *STOC '98 (Dallas, TX)*. ACM, New York, 1998, pp. 644–652.

[MS15] T. Maciążek and A. Sawicki. “Critical points of the linear entropy for pure L-qubit states”. In: *J. Phys. A* 48.4 (2015), pp. 045305, 25. doi: 10.1088/1751-8113/48/4/045305.

[MS18] T. Maciążek and A. Sawicki. “Asymptotic properties of entanglement polytopes for large number of qubits”. In: *J. Phys. A* 51.7 (2018), 07LT01, 11. doi: 10.1088/1751-8121/aaa4d7.

[Rus20] A. Rusciano. “A Riemannian Corollary of Helly's theorem”. In: *J. Convex Anal.* 27.4 (2020), pp. 1261–1275. issn: 0944-6532.

References IV

[Sja98] R. Sjamaar. “Convexity properties of the moment mapping re-examined”. In: *Adv. Math.* 138.1 (1998), pp. 46–91. doi: 10.1006/aima.1998.1739.

[SV19] D. Straszak and N. K. Vishnoi. “Maximum Entropy Distributions: Bit Complexity and Stability”. In: *Proceedings of the Thirty-Second Conference on Learning Theory*. Vol. 99. Proceedings of Machine Learning Research. PMLR, 25–28 Jun 2019, pp. 2861–2891. arXiv: 1711.02036. url: https://proceedings.mlr.press/v99/straszak19a.html.

Please find the full list of references in our paper!

Polynomial Scaling

Consider the scaling action on homogeneous polynomials of degree d:

$$\varrho_{n,d} \colon \ \mathrm{SL}(n) \to \mathrm{GL}\left(\mathbb{C}[x_1, \ldots, x_n]_d\right), \ g \mapsto \left(p(x) \mapsto p(g^{-1} \cdot x)\right)$$

As a consequence of the theorem for d-tensor scaling we obtain

Theorem: Gap for polynomial scaling [Franks, R.]

There exists a constant $C > 0$ such that for $d \geq 3$ and $n = dm$, where $m \geq 2$, one has

$$\gamma_{\mathrm{T}}(\varrho_{n,d}) \leq \gamma_G(\varrho_{n,d}) \leq 2^{-Cdm} = 2^{-Cn}.$$

More concretely, for $d = 3$ (i.e., *cubic forms*) and $m \geq 3$ it holds that

$$\gamma_{\mathrm{T}}(\varrho_{n,3}) \leq \gamma_G(\varrho_{n,3}) \leq 2^{-m+1} = 2^{-\frac{n}{3}+1}.$$

Bound on $\gamma_{\mathrm{T}}(\pi_{n,3})$: the witness set $\Gamma_{n,3}$

Recall: $\Omega(\pi_{n,3}) = \big\{(\varepsilon_i, \varepsilon_j, \varepsilon_k) \mid i, j, k \in [n]\big\}$

$\gamma_{\mathrm{T}}(\pi_{n,3}) = \min \big\{ \operatorname{dist}\big(0, \operatorname{conv}(S)\big) \mid S \subseteq \Omega(\pi_{n,3}),\ 0 \notin \operatorname{conv}(S) \big\}$

Want: $\Gamma_{n,3} \subseteq \Omega(\pi_{n,3})$ with $0 < \operatorname{dist}\big(0, \operatorname{conv}(\Gamma_{n,3})\big) \leq 2^{-n+1}$

Bound on $\gamma_{\mathrm{T}}(\pi_{n,3})$: the witness set $\Gamma_{n,3}$

Recall: $\Omega(\pi_{n,3}) = \{(\varepsilon_i, \varepsilon_j, \varepsilon_k) \mid i, j, k \in [n]\}$

$\gamma_{\mathrm{T}}(\pi_{n,3}) = \min\{\operatorname{dist}(0, \operatorname{conv}(S)) \mid S \subseteq \Omega(\pi_{n,3}),\ 0 \notin \operatorname{conv}(S)\}$

Want: $\Gamma_{n,3} \subseteq \Omega(\pi_{n,3})$ with $0 < \operatorname{dist}(0, \operatorname{conv}(\Gamma_{n,3})) \leq 2^{-n+1}$

- [Kra07]: the *"3-dimensional Birkhoff polytope"*

  $B_{n,3} = \{(x_{ijk}) \in (\mathbb{R}^n_{\geq 0})^{\otimes 3} \mid \forall i_0 \in [n]\colon\ \sum_{j,k} x_{i_0,j,k} = 1 \text{ etc.}\}$

  has a vertex (λ_{ijk}) with $\lambda_{111} = 2^{-n+1}$, support $\{(1,1,1)\} \cup \mathfrak{W}_n$, where $\mathfrak{W}_n = \{(s,1,s), (s,s,1), (s-1,s,s) \mid s = 2, \ldots, n\}$.

Bound on $\gamma_{\mathrm{T}}(\pi_{n,3})$: the witness set $\Gamma_{n,3}$

Recall: $\Omega(\pi_{n,3}) = \{(\varepsilon_i, \varepsilon_j, \varepsilon_k) \mid i,j,k \in [n]\}$

$\gamma_{\mathrm{T}}(\pi_{n,3}) = \min\{\, \mathrm{dist}\,(0, \mathrm{conv}(S)) \mid S \subseteq \Omega(\pi_{n,3}),\ 0 \notin \mathrm{conv}(S)\}$

Want: $\Gamma_{n,3} \subseteq \Omega(\pi_{n,3})$ with $0 < \mathrm{dist}\,(0, \mathrm{conv}(\Gamma_{n,3})) \leq 2^{-n+1}$

- [Kra07]: the *"3-dimensional Birkhoff polytope"*

  $B_{n,3} = \{(x_{ijk}) \in (\mathbb{R}^n_{\geq 0})^{\otimes 3} \mid \forall i_0 \in [n]\colon\ \sum_{j,k} x_{i_0,j,k} = 1 \text{ etc.}\}$

  has a vertex (λ_{ijk}) with $\lambda_{111} = 2^{-n+1}$, support $\{(1,1,1)\} \cup \mathfrak{W}_n$, where $\mathfrak{W}_n = \{(s,1,s), (s,s,1), (s-1,s,s) \mid s = 2, \dots, n\}$.

- choose $\Gamma_{n,3} := \{(\varepsilon_i, \varepsilon_j, \varepsilon_k) \mid (i,j,k) \in \mathfrak{W}_n\}$

- $\sum_i \varepsilon_i = 0 \quad \Rightarrow \quad 0 =$
  $2^{n-1}(\varepsilon_1, \varepsilon_1, \varepsilon_1) + \sum_{(i,j,k) \in \mathfrak{W}_n} \lambda_{ijk}(\varepsilon_i, \varepsilon_j, \varepsilon_k)$

  $\Rightarrow \mathrm{dist}\,(0, \mathrm{conv}(\Gamma_{n,3})) \leq 2^{-n+1}$

Proof of $0 \notin \mathrm{conv}(\Gamma_{n,3})$

$(*)$ $\sum_i \frac{1}{n} \varepsilon_i$ is the only convex combination of $\varepsilon_1, \ldots, \varepsilon_n$ giving zero.

- Assume $0 \in \mathrm{conv}(\Gamma_{n,3})$. Then there are $a_s, b_s, c_s \geq 0$ with
$$\sum_{s=2}^{n} \Big(a_s(\varepsilon_s, \varepsilon_1, \varepsilon_s) + b_s(\varepsilon_s, \varepsilon_s, \varepsilon_1) + c_s(\varepsilon_{s-1}, \varepsilon_s, \varepsilon_s) \Big) = 0 \in (\mathbb{R}^n)^3$$

- For $s = 2, \ldots, n$, apply $(*)$ to coeff. of ε_{s-1} in first component resp. to coeff. of ε_s in second, third component:

  (I) $a_{s-1} + b_{s-1} + c_s = n^{-1}$ resp. (II) $b_s + c_s = n^{-1} = a_s + c_s$

Proof of $0 \notin \operatorname{conv}(\Gamma_{n,3})$

$(*)$ $\sum_i \frac{1}{n} \varepsilon_i$ is the only convex combination of $\varepsilon_1, \ldots, \varepsilon_n$ giving zero.

- Assume $0 \in \operatorname{conv}(\Gamma_{n,3})$. Then there are $a_s, b_s, c_s \geq 0$ with

$$\sum_{s=2}^{n} \Big(a_s(\varepsilon_s, \varepsilon_1, \varepsilon_s) + b_s(\varepsilon_s, \varepsilon_s, \varepsilon_1) + c_s(\varepsilon_{s-1}, \varepsilon_s, \varepsilon_s) \Big) = 0 \in (\mathbb{R}^n)^3$$

- For $s = 2, \ldots, n$, apply $(*)$ to coeff. of ε_{s-1} in first component resp. to coeff. of ε_s in second, third component:

  (I) $a_{s-1} + b_{s-1} + c_s = n^{-1}$ resp. (II) $b_s + c_s = n^{-1} = a_s + c_s$

- (I) shows $c_2 = n^{-1}$; (II) gives $a_2 = b_2 = 0$;

Proof of $0 \notin \mathrm{conv}(\Gamma_{n,3})$

$(*)$ $\sum_i \frac{1}{n} \varepsilon_i$ is the only convex combination of $\varepsilon_1, \dots, \varepsilon_n$ giving zero.

- Assume $0 \in \mathrm{conv}(\Gamma_{n,3})$. Then there are $a_s, b_s, c_s \geq 0$ with
$$\sum_{s=2}^{n} \Big(a_s(\varepsilon_s, \varepsilon_1, \varepsilon_s) + b_s(\varepsilon_s, \varepsilon_s, \varepsilon_1) + c_s(\varepsilon_{s-1}, \varepsilon_s, \varepsilon_s) \Big) = 0 \in (\mathbb{R}^n)^3$$

- For $s = 2, \dots, n$, apply $(*)$ to coeff. of ε_{s-1} in first component resp. to coeff. of ε_s in second, third component:
(I) $a_{s-1} + b_{s-1} + c_s = n^{-1}$ resp. (II) $b_s + c_s = n^{-1} = a_s + c_s$

- (I) shows $c_2 = n^{-1}$; (II) gives $a_2 = b_2 = 0$; (I) yields $c_3 = n^{-1}$

Proof of $0 \notin \mathrm{conv}(\Gamma_{n,3})$

$(*)$ $\sum_i \frac{1}{n}\varepsilon_i$ is the only convex combination of $\varepsilon_1, \ldots, \varepsilon_n$ giving zero.

- Assume $0 \in \mathrm{conv}(\Gamma_{n,3})$. Then there are $a_s, b_s, c_s \geq 0$ with
$$\sum_{s=2}^{n} \Big(a_s(\varepsilon_s, \varepsilon_1, \varepsilon_s) + b_s(\varepsilon_s, \varepsilon_s, \varepsilon_1) + c_s(\varepsilon_{s-1}, \varepsilon_s, \varepsilon_s) \Big) = 0 \in (\mathbb{R}^n)^3$$
- For $s = 2, \ldots, n$, apply $(*)$ to coeff. of ε_{s-1} in first component resp. to coeff. of ε_s in second, third component:
(I) $a_{s-1} + b_{s-1} + c_s = n^{-1}$ resp. (II) $b_s + c_s = n^{-1} = a_s + c_s$
- (I) shows $c_2 = n^{-1}$; (II) gives $a_2 = b_2 = 0$; (I) yields $c_3 = n^{-1}$
- inductively: $a_s = b_s = 0$ and $c_s = n^{-1}$ for $s = 2, \ldots, n$ $\lightning$