Non-Commutative Optimization

Peter Bürgisser

IPAM workshop IV: Efficient Tensor Representations for Learning and Computational Complexity

May 20, 2021

based on joint work with

Cole Franks, Ankit Garg, Rafael Oliveira, Michael Walter, Avi Wigderson

ITCS 2018, FOCS 2018, FOCS 2019

Full version at arXiv:1910.12375, ca. 100 pages

Technische Universität Berlin
Berlin Mathematics Research Center MATH+
European Research Council
Established by the European Commission
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

General setting

Tensor Action (main example):
Action of group $G = \mathrm{SL}_{m_1} \times \cdots \times \mathrm{SL}_{m_p}$ on space of tensors $V = \mathbb{C}^{m_1} \otimes \cdots \otimes \mathbb{C}^{m_p}$ via tensor product

Baby example: conj. action of $G = \mathrm{GL}_n$ on $V = \mathbb{C}^{n \times n}$, $\pi(g)(A) = gAg^{-1}$

General setting:

- Algebraic subgroup $G \leq \mathrm{GL}_n$ closed under $g \mapsto g^*$ (complex reductive group), e.g., $T_n := (\mathbb{C}^*)^n$, GL_n, SL_n or products
- Regular representation $\pi: G \to \mathrm{GL}(V)$ on f.d. $\mathbb{C}$-Hilbert space V: G acts on V by linear transformations
- Maximal compact subgroup $K = G \cap U_n$ of G
- Assume K acts isometrically on V
- Orbit $Gv := \{\pi(g)v \mid g \in G\}$ of $v \in V$ has closure $\overline{Gv}$

General setting

Tensor Action (main example):
Action of group $G = \mathrm{SL}_{m_1} \times \cdots \times \mathrm{SL}_{m_p}$ on space of tensors $V = \mathbb{C}^{m_1} \otimes \cdots \otimes \mathbb{C}^{m_p}$ via tensor product

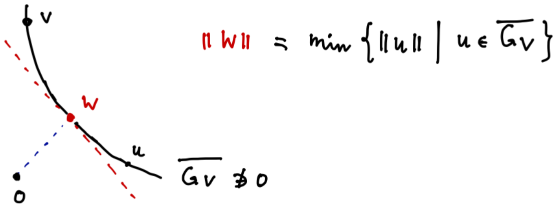
Baby example: conj. action of $G = \mathrm{GL}_n$ on $V = \mathbb{C}^{n \times n}$, $\pi(g)(A) = gAg^{-1}$

General setting:

- Algebraic subgroup $G \leq \mathrm{GL}_n$ closed under $g \mapsto g^*$ (complex reductive group), e.g., $T_n := (\mathbb{C}^*)^n$, GL_n, SL_n or products
- Regular representation $\pi \colon G \to \mathrm{GL}(V)$ on f.d. $\mathbb{C}$-Hilbert space V: G acts on V by linear transformations
- Maximal compact subgroup $K = G \cap U_n$ of G
- Assume K acts isometrically on V
- Orbit $Gv := \{\pi(g)v \mid g \in G\}$ of $v \in V$ has closure $\overline{Gv}$

Norm minimization

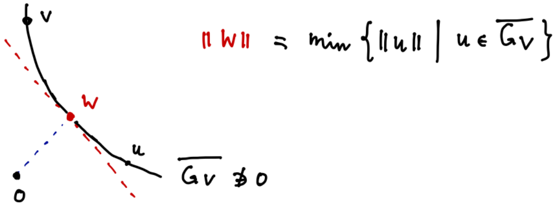
- For given $v \in V$ (approximately) compute $g \in G$ such that $\pi(g)v$ has minimal norm:



- Capacity of v: $\mathrm{cap}(v) := \inf_{g \in G} \|\pi(g)v\|$
- Conj. of matrices:
  diagonal matrix w with same eigenvalues as v minimizes norm
- Shortest vector w is uniquely determined up to unitary action (Kempf-Ness 1978)

Norm minimization

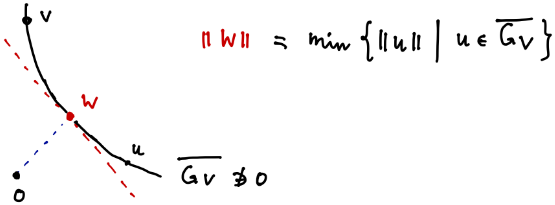
- For given $v \in V$ (approximately) compute $g \in G$ such that $\pi(g)v$ has minimal norm:



- Capacity of v: $\mathrm{cap}(v) := \inf_{g \in G} \|\pi(g)v\|$
- Conj. of matrices:
  diagonal matrix w with same eigenvalues as v minimizes norm
- Shortest vector w is uniquely determined up to unitary action (Kempf-Ness 1978)

Norm minimization

- For given $v \in V$ (approximately) compute $g \in G$ such that $\pi(g)v$ has minimal norm:



- Capacity of v: $\mathrm{cap}(v) := \inf_{g \in G} \|\pi(g)v\|$
- Conj. of matrices:
  diagonal matrix w with same eigenvalues as v minimizes norm
- Shortest vector w is uniquely determined up to unitary action (Kempf-Ness 1978)

Geodesic convexity

- Use techniques from convex optimization on Riemannian manifolds
- Riemannian geometry induced by Lie group symmetry
- For simplicity: $G = \mathrm{GL}_n, K = U_n$
- Want to minimize $F_v \colon G \to \mathbb{R},\ g \mapsto \log \|\pi(g)v\|$

$$
\begin{tikzcd}
g \in GL_n \arrow[r, "F"] \arrow[d] & \mathbb{R} \ni \log\|\pi(g)v\| \\
X = g^* g \in PSD_n \arrow[ur, "\tilde{F}"] &
\end{tikzcd}
$$

$$= \tfrac{1}{2}\log\langle \pi(X)v, v\rangle$$

- Cone PSD_n of hermitian psd matrices has natural Riemannian metric (invariant under conjugation)
- Induced map $\tilde{F}_v \colon \mathrm{PSD}_n \to \mathbb{R},\ X \mapsto \frac{1}{2}\log\langle \pi(X)v, v\rangle$ is geodesically convex (implicit in Kemp-Ness 1978)

Geodesic convexity

- Use techniques from convex optimization on Riemannian manifolds
- Riemannian geometry induced by Lie group symmetry
- For simplicity: $G = \mathrm{GL}_n, K = U_n$
- Want to minimize $F_v \colon G \to \mathbb{R},\ g \mapsto \log \|\pi(g) v\|$

$$
\begin{tikzcd}
g \in GL_n \arrow[r, "F"] \arrow[d] & \mathbb{R} \ni \log\|\pi(g)v\| \\
X = g^*g \in PSD_n \arrow[ur, "\tilde{F}"] & = \frac{1}{2}\log\langle \pi(X)v, v\rangle
\end{tikzcd}
$$

- Cone PSD_n of hermitian psd matrices has natural Riemannian metric (invariant under conjugation)
- Induced map $\tilde{F}_v \colon \mathrm{PSD}_n \to \mathbb{R},\ X \mapsto \frac{1}{2}\log\langle \pi(X) v, v\rangle$ is geodesically convex (implicit in Kemp-Ness 1978)

Dual norm minimization (or scaling)

- We want to minimize

$$F_v \colon G \to \mathbb{R},\ g \mapsto \log \|\pi(g)v\|$$

- Consider derivative or gradient $\mu(v) \in \mathrm{Lie}(G) \subseteq \mathbb{C}^{n\times n}$ at $g_0 = I$ of F_v
- Conj. of matrices: $\mu(A) = \|A\|_F^{-2}(AA^* - A^*A)$
- Tensor action: for $v \in \mathbb{C}^{m_1} \otimes \cdots \otimes \mathbb{C}^{m_p}$, have $\mu(v) = (\rho_1, \ldots, \rho_p)$, where $\rho_k = \mathrm{tr}_k(vv^*) \in \mathbb{C}^{m_i \times m_i}$ is kth partial trace, e.g.,

$$(\rho_1)_{ij} = \sum_{i_2,\ldots,i_p} v_{ii_2\ldots i_p} \bar{v}_{ji_2\ldots i_p}.$$

- Dual norm minimization:
  For given $v \in V$ (approximately) compute $g \in G$ such that $\|\mu(\pi(g)v)\|_F$ has minimal norm
- Kempf-Ness: Suppose $\mathrm{cap}(v) > 0$. Then

$$w \in \overline{Gv} \text{ has minimal norm iff } \mu(w) = 0$$

Dual norm minimization (or scaling)

- We want to minimize

$$F_v \colon G \to \mathbb{R},\ g \mapsto \log \|\pi(g)v\|$$

- Consider derivative or gradient $\mu(v) \in \mathrm{Lie}(G) \subseteq \mathbb{C}^{n\times n}$ at $g_0 = I$ of F_v
- Conj. of matrices: $\mu(A) = \|A\|_F^{-2}(AA^* - A^*A)$
- Tensor action: for $v \in \mathbb{C}^{m_1} \otimes \cdots \otimes \mathbb{C}^{m_p}$, have $\mu(v) = (\rho_1, \ldots, \rho_p)$, where $\rho_k = \mathrm{tr}_k(vv^*) \in \mathbb{C}^{m_i \times m_i}$ is kth partial trace, e.g.,

$$(\rho_1)_{ij} = \sum_{i_2,\ldots,i_p} v_{ii_2\ldots i_p} \bar{v}_{ji_2\ldots i_p}.$$

- Dual norm minimization:
  For given $v \in V$ (approximately) compute $g \in G$ such that $\|\mu(\pi(g)v)\|_F$ has minimal norm
- Kempf-Ness: Suppose $\mathrm{cap}(v) > 0$. Then

$$w \in \overline{Gv} \text{ has minimal norm iff } \mu(w) = 0$$

Dual norm minimization (or scaling)

- We want to minimize

$$F_v : G \to \mathbb{R},\ g \mapsto \log \|\pi(g)v\|$$

- Consider derivative or gradient $\mu(v) \in \mathrm{Lie}(G) \subseteq \mathbb{C}^{n\times n}$ at $g_0 = I$ of F_v
- Conj. of matrices: $\mu(A) = \|A\|_F^{-2}(AA^* - A^*A)$
- Tensor action: for $v \in \mathbb{C}^{m_1} \otimes \cdots \otimes \mathbb{C}^{m_p}$, have $\mu(v) = (\rho_1, \ldots, \rho_p)$, where $\rho_k = \mathrm{tr}_k(vv^*) \in \mathbb{C}^{m_i \times m_i}$ is kth partial trace, e.g.,

$$(\rho_1)_{ij} = \sum_{i_2,\ldots,i_p} v_{ii_2\ldots i_p} \bar{v}_{ji_2\ldots i_p}.$$

- Dual norm minimization:
  For given $v \in V$ (approximately) compute $g \in G$ such that $\|\mu(\pi(g)v)\|_F$ has minimal norm
- Kempf-Ness: Suppose $\mathrm{cap}(v) > 0$. Then

$$w \in \overline{Gv} \text{ has minimal norm iff } \mu(w) = 0$$

Dual norm minimization (or scaling)

- We want to minimize

$$F_v \colon G \to \mathbb{R},\ g \mapsto \log \|\pi(g)v\|$$

- Consider derivative or gradient $\mu(v) \in \mathrm{Lie}(G) \subseteq \mathbb{C}^{n \times n}$ at $g_0 = I$ of F_v
- Conj. of matrices: $\mu(A) = \|A\|_F^{-2}(AA^* - A^*A)$
- Tensor action: for $v \in \mathbb{C}^{m_1} \otimes \cdots \otimes \mathbb{C}^{m_p}$, have $\mu(v) = (\rho_1, \ldots, \rho_p)$, where $\rho_k = \mathrm{tr}_k(vv^*) \in \mathbb{C}^{m_i \times m_i}$ is kth partial trace, e.g.,

$$(\rho_1)_{ij} = \sum_{i_2,\ldots,i_p} v_{ii_2\ldots i_p} \bar{v}_{ji_2\ldots i_p}.$$

- Dual norm minimization:
  For given $v \in V$ (approximately) compute $g \in G$ such that $\|\mu(\pi(g)v)\|_F$ has minimal norm
- Kempf-Ness: Suppose $\mathrm{cap}(v) > 0$. Then

$$w \in \overline{Gv} \text{ has minimal norm iff } \mu(w) = 0$$

Dual norm minimization (or scaling)

- We want to minimize

$$F_v \colon G \to \mathbb{R},\ g \mapsto \log \|\pi(g) v\|$$

- Consider derivative or gradient $\mu(v) \in \mathrm{Lie}(G) \subseteq \mathbb{C}^{n\times n}$ at $g_0 = I$ of F_v
- Conj. of matrices: $\mu(A) = \|A\|_F^{-2}(AA^* - A^* A)$
- Tensor action: for $v \in \mathbb{C}^{m_1} \otimes \cdots \otimes \mathbb{C}^{m_p}$, have $\mu(v) = (\rho_1, \ldots, \rho_p)$, where $\rho_k = \mathrm{tr}_k(vv^*) \in \mathbb{C}^{m_i \times m_i}$ is kth partial trace, e.g.,

$$(\rho_1)_{ij} = \sum_{i_2,\ldots,i_p} v_{ii_2\ldots i_p} \bar{v}_{ji_2\ldots i_p}.$$

- Dual norm minimization:
  For given $v \in V$ (approximately) compute $g \in G$ such that $\|\mu(\pi(g)v)\|_F$ has minimal norm
- Kempf-Ness: Suppose $\mathrm{cap}(v) > 0$. Then

$$w \in \overline{Gv} \text{ has minimal norm iff } \mu(w) = 0$$

First order algorithm for scaling

- Developed gradient descent algorithm: given $v \in V$ such that $\mathrm{cap}(v) > 0$, it outputs $g \in G$ such that $\|\mu(\pi(g)v)\|_F \leq \varepsilon$, with number of iterations

$$\frac{4N(\pi)^2}{\varepsilon^2} \log \frac{\|v\|}{\mathrm{cap}(v)}$$

- Hilbert associated with a representation π its null cone

$$\mathcal{N}(\pi) := \{v \in V \mid \mathrm{cap}(v) = 0\} = \{v \in V \mid 0 \in \overline{Gv}\}$$

Conj. of matrices: $\mathcal{N}(\pi)$ consists of nilpotent matrices

- Can use above algorithm for testing membership to the null cone.
- This is a special case of moment polytope membership problem, of relevance in quantum information theory.
- Developed and analyzed gradient descent algorithm for p-scaling and deciding (approximate) membership to moment polytopes.

First order algorithm for scaling

- Developed gradient descent algorithm: given $v \in V$ such that $\mathrm{cap}(v) > 0$, it outputs $g \in G$ such that $\|\mu(\pi(g)v)\|_F \leq \varepsilon$, with number of iterations

$$\frac{4N(\pi)^2}{\varepsilon^2} \log \frac{\|v\|}{\mathrm{cap}(v)}$$

- Hilbert associated with a representation π its null cone

$$\mathcal{N}(\pi) := \{v \in V \mid \mathrm{cap}(v) = 0\} = \{v \in V \mid 0 \in \overline{Gv}\}$$

Conj. of matrices: $\mathcal{N}(\pi)$ consists of nilpotent matrices

- Can use above algorithm for testing membership to the null cone.
- This is a special case of moment polytope membership problem, of relevance in quantum information theory.
- Developed and analyzed gradient descent algorithm for p-scaling and deciding (approximate) membership to moment polytopes.

First order algorithm for scaling

- Developed gradient descent algorithm: given $v \in V$ such that $\operatorname{cap}(v) > 0$, it outputs $g \in G$ such that $\|\mu(\pi(g)v)\|_F \leq \varepsilon$, with number of iterations

$$\frac{4N(\pi)^2}{\varepsilon^2} \log \frac{\|v\|}{\operatorname{cap}(v)}$$

- Hilbert associated with a representation π its null cone

$$\mathscr{N}(\pi) := \{v \in V \mid \operatorname{cap}(v) = 0\} = \{v \in V \mid 0 \in \overline{Gv}\}$$

Conj. of matrices: $\mathscr{N}(\pi)$ consists of nilpotent matrices

- Can use above algorithm for testing membership to the null cone.
- This is a special case of moment polytope membership problem, of relevance in quantum information theory.
- Developed and analyzed gradient descent algorithm for p-scaling and deciding (approximate) membership to moment polytopes.

First order algorithm for scaling

- Developed gradient descent algorithm: given $v \in V$ such that $\mathrm{cap}(v) > 0$, it outputs $g \in G$ such that $\|\mu(\pi(g)v)\|_F \leq \varepsilon$, with number of iterations

$$\frac{4N(\pi)^2}{\varepsilon^2} \log \frac{\|v\|}{\mathrm{cap}(v)}$$

- Hilbert associated with a representation π its null cone

$$\mathcal{N}(\pi) := \{v \in V \mid \mathrm{cap}(v) = 0\} = \{v \in V \mid 0 \in \overline{Gv}\}$$

  Conj. of matrices: $\mathcal{N}(\pi)$ consists of nilpotent matrices
- Can use above algorithm for testing membership to the null cone.
- This is a special case of moment polytope membership problem, of relevance in quantum information theory.
- Developed and analyzed gradient descent algorithm for p-scaling and deciding (approximate) membership to moment polytopes.

Algorithm for deciding null cone membership

- We shall identify gap parameters $\gamma(\pi)$ and $c(\pi)$ such that

$$\forall v \in V \setminus 0 \quad \operatorname{cap}(v) = 0 \Longrightarrow \inf_{g \in G} \|\mu(\pi(g)v)\|_F \geq \gamma(\pi)$$

$$\forall v \in V(\mathbb{Z}[i]) \quad \operatorname{cap}(v) > 0 \Longrightarrow \operatorname{cap}(v) \geq c(\pi)$$

- Run gradient descent on v for $T = \frac{16N(\pi)}{\gamma(\pi)^2} \log\left(\frac{\|v\|}{c(\pi)}\right)$ iterations, producing $g_1, \ldots, g_T$.
- If $\|\mu(\pi(g_t)v)\|_F \leq \frac{1}{2}\gamma(\pi)$ for some $t \leq T$, then output "$\operatorname{cap}(v) > 0$". Otherwise, output "$\operatorname{cap}(v) = 0$".

Algorithm for deciding null cone membership

- We shall identify gap parameters $\gamma(\pi)$ and $c(\pi)$ such that

$$\forall v \in V \setminus 0 \quad \mathrm{cap}(v) = 0 \Longrightarrow \inf_{g \in G} \|\mu(\pi(g)v)\|_F \geq \gamma(\pi)$$

$$\forall v \in V(\mathbb{Z}[i]) \quad \mathrm{cap}(v) > 0 \Longrightarrow \mathrm{cap}(v) \geq c(\pi)$$

- Run gradient descent on v for $T = \frac{16N(\pi)}{\gamma(\pi)^2} \log\left(\frac{\|v\|}{c(\pi)}\right)$ iterations, producing $g_1, \ldots, g_T$.
- If $\|\mu(\pi(g_t)v)\|_F \leq \frac{1}{2}\gamma(\pi)$ for some $t \leq T$, then output "$\mathrm{cap}(v) > 0$". Otherwise, output "$\mathrm{cap}(v) = 0$".

Critical complexity parameter

- Gradient descent decides membership to null cone with # iterations

$$T = \frac{16N(\pi)}{\gamma(\pi)^2} \log\left(\frac{\|v\|}{c(\pi)}\right)$$

- # iterations is poly bounded in weight norm $N(\pi) \leq \mathrm{degree}$, $\log\|v\|$, $\log c(\pi)^{-1}$, and $\gamma(\pi)^{-1}$.
- **Thm.** $\log c(\pi)^{-1} \leq \mathrm{poly}(m, n, d)$ for homogeneous polynomial representations $\pi : \mathrm{SL}(n) \to \mathrm{GL}(m)$ of degree d, e.g., when π given in Gelfand-Tsetlin basis.
- Proof based on deep general results from invariant theory: (1) Derksen's general degree bound for a system of generators of ring of invariants. (2) Analysis of Cayley's Ω-process for SL_n.
- So $c(\pi)$ is harmless, $\gamma(\pi)$ is the critical parameter.

Critical complexity parameter

- Gradient descent decides membership to null cone with # iterations

$$T = \frac{16N(\pi)}{\gamma(\pi)^2} \log\left(\frac{\|v\|}{c(\pi)}\right)$$

- # iterations is poly bounded in weight norm $N(\pi) \le \mathrm{degree}$, $\log\|v\|$, $\log c(\pi)^{-1}$, and $\gamma(\pi)^{-1}$.
- **Thm.** $\log c(\pi)^{-1} \le \mathrm{poly}(m, n, d)$ for homogeneous polynomial representations $\pi\colon \mathrm{SL}(n) \to \mathrm{GL}(m)$ of degree d, e.g., when π given in Gelfand-Tsetlin basis.
- Proof based on deep general results from invariant theory: (1) Derksen's general degree bound for a system of generators of ring of invariants. (2) Analysis of Cayley's Ω-process for SL_n.
- So $c(\pi)$ is harmless, $\gamma(\pi)$ is the critical parameter.

Critical complexity parameter

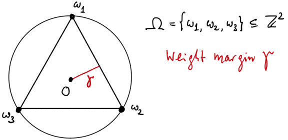
- Gradient descent decides membership to null cone with # iterations

$$T = \frac{16N(\pi)}{\gamma(\pi)^2} \log\left(\frac{\|v\|}{c(\pi)}\right)$$

- # iterations is poly bounded in weight norm $N(\pi) \leq \mathrm{degree}$, $\log\|v\|$, $\log c(\pi)^{-1}$, and $\gamma(\pi)^{-1}$.
- **Thm.** $\log c(\pi)^{-1} \leq \mathrm{poly}(m, n, d)$ for homogeneous polynomial representations $\pi\colon \mathrm{SL}(n) \to \mathrm{GL}(m)$ of degree d, e.g., when π given in Gelfand-Tsetlin basis.
- Proof based on deep general results from invariant theory: (1) Derksen's general degree bound for a system of generators of ring of invariants. (2) Analysis of Cayley's Ω-process for SL_n.
- So $c(\pi)$ is harmless, $\gamma(\pi)$ is the critical parameter.

Weight margin

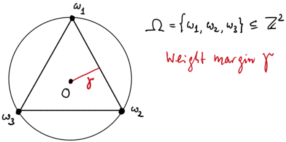
- Can cleanly identify a dual gap parameter $\gamma(\pi)$
- For simplicity assume $G = \mathrm{GL}_n$ or $G = T_n$
- Let $\Omega(\pi) \subseteq \mathbb{Z}^n$ be the set of weights of π
- Define the weight margin $\gamma(\pi)$ as the minimum distance between 0 and the convex hull of any subset of $\Omega(\pi)$ which does not contain 0



- **Prop.** The weight margin $\gamma(\pi)$ is a dual gap parameter.
- **Thm.** GL-representations of quivers (e.g. L/R action) have inverse polynomial lower bounds for $\gamma(\pi)$. (Key insight: totally unimodular matrix of weights.) Scaling algorithm for null cone is poly time.
- For tensor action $\gamma(\pi)$ can get exponentially small [Kravtsov, Franks & Reichenbach]
- More sophisticated analysis in [Franks & Reichenbach, CCC 21]

Weight margin

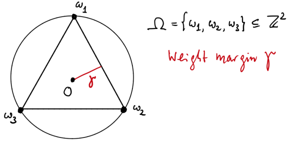
- Can cleanly identify a dual gap parameter $\gamma(\pi)$
- For simplicity assume $G = \mathrm{GL}_n$ or $G = T_n$
- Let $\Omega(\pi) \subseteq \mathbb{Z}^n$ be the set of weights of π
- Define the weight margin $\gamma(\pi)$ as the minimum distance between 0 and the convex hull of any subset of $\Omega(\pi)$ which does not contain 0



- **Prop.** The weight margin $\gamma(\pi)$ is a dual gap parameter.
- **Thm.** GL-representations of quivers (e.g. L/R action) have inverse polynomial lower bounds for $\gamma(\pi)$. (Key insight: totally unimodular matrix of weights.) Scaling algorithm for null cone is poly time.
- For tensor action $\gamma(\pi)$ can get exponentially small [Kravtsov, Franks & Reichenbach]
- More sophisticated analysis in [Franks & Reichenbach, CCC 21]

Weight margin

- Can cleanly identify a dual gap parameter $\gamma(\pi)$
- For simplicity assume $G = \mathrm{GL}_n$ or $G = T_n$
- Let $\Omega(\pi) \subseteq \mathbb{Z}^n$ be the set of weights of π
- Define the weight margin $\gamma(\pi)$ as the minimum distance between 0 and the convex hull of any subset of $\Omega(\pi)$ which does not contain 0



- **Prop.** The weight margin $\gamma(\pi)$ is a dual gap parameter.
- **Thm.** GL-representations of quivers (e.g. L/R action) have inverse polynomial lower bounds for $\gamma(\pi)$. (Key insight: totally unimodular matrix of weights.) Scaling algorithm for null cone is poly time.
- For tensor action $\gamma(\pi)$ can get exponentially small [Kravtsov, Franks & Reichenbach]
- More sophisticated analysis in [Franks & Reichenbach, CCC 21]

Quantitative duality

Thm.

For $v \in V \setminus 0$ we have

$$1 - \frac{\|\mu(v)\|_F}{\gamma(\pi)} \le \frac{\mathrm{cap}(v)^2}{\|v\|^2} \le 1 - \frac{\|\mu(v)\|_F^2}{4N(\pi)^2}$$

This again shows "$\mu(v) = 0 \Rightarrow \|v\| = \mathrm{cap}(v)$", but turns this into a quantitative statement.

Another consequence: weight margin indeed gives dual gap. If $\mathrm{cap}(v) = 0$, then

$$\gamma(\pi) \le \min_{g \in G} \|\mu(gv)\|_F$$

Quantitative duality

Thm.

For $v \in V \setminus 0$ we have

$$1 - \frac{\|\mu(v)\|_F}{\gamma(\pi)} \leq \frac{\mathrm{cap}(v)^2}{\|v\|^2} \leq 1 - \frac{\|\mu(v)\|_F^2}{4N(\pi)^2}$$

This again shows "$\mu(v) = 0 \Rightarrow \|v\| = \mathrm{cap}(v)$", but turns this into a quantitative statement.

Another consequence: weight margin indeed gives dual gap. If $\mathrm{cap}(v) = 0$, then

$$\gamma(\pi) \leq \min_{g \in G} \|\mu(gv)\|_F$$

Quantitative duality

Thm.

For $v \in V \setminus 0$ we have

$$1 - \frac{\|\mu(v)\|_F}{\gamma(\pi)} \leq \frac{\mathrm{cap}(v)^2}{\|v\|^2} \leq 1 - \frac{\|\mu(v)\|_F^2}{4N(\pi)^2}$$

This again shows "$\mu(v) = 0 \Rightarrow \|v\| = \mathrm{cap}(v)$", but turns this into a quantitative statement.

Another consequence: weight margin indeed gives dual gap. If $\mathrm{cap}(v) = 0$, then

$$\gamma(\pi) \leq \min_{g \in G} \|\mu(gv)\|_F$$

Null cone membership: previously known results

- Simultaneous LR-action: $G = \mathrm{SL}_n \times \mathrm{SL}_n$ acts on $V = (\mathbb{C}^{n \times n})^m$:

$$\pi(g, h)(A_1, \ldots, A_m) := (gA_1h^T, \ldots, gA_mh^T)$$

- Alternating minimization algorithm by [Gurvits 2004)], generalizing Sinkhorn's algorithm, shown to be poly time in [Garg, Gurvits, Oliveira, Wigderson 2016] using invariant theory
- Algebraic poly time algorithms [Ivanyos, Qiao, Subrahmanyam 2017] works in any char
- Tensor action: Alternating minimization algorithm by [Verstrate et al 2004] analyzed by [B, Franks, Garg, Oliveira, Walter, Wigderson 2018]. However, only get exponential time algorithm!

Are there poly time algorithms for null cone memberships for tensors?

Null cone membership: previously known results

- Simultaneous LR-action: $G = \mathrm{SL}_n \times \mathrm{SL}_n$ acts on $V = (\mathbb{C}^{n\times n})^m$:

$$\pi(g,h)(A_1,\ldots,A_m) := (gA_1h^T,\ldots,gA_mh^T)$$

- Alternating minimization algorithm by [Gurvits 2004)], generalizing Sinkhorn's algorithm, shown to be poly time in [Garg, Gurvits, Oliveira, Wigderson 2016] using invariant theory
- Algebraic poly time algorithms [Ivanyos, Qiao, Subrahmanyam 2017] works in any char
- Tensor action: Alternating minimization algorithm by [Verstrate et al 2004] analyzed by [B, Franks, Garg, Oliveira, Walter, Wigderson 2018]. However, only get exponential time algorithm!

Are there poly time algorithms for null cone memberships for tensors?

Null cone membership: previously known results

- Simultaneous LR-action: $G = \mathrm{SL}_n \times \mathrm{SL}_n$ acts on $V = (\mathbb{C}^{n\times n})^m$:

$$\pi(g,h)(A_1,\ldots,A_m) := (gA_1h^T,\ldots,gA_mh^T)$$

- Alternating minimization algorithm by [Gurvits 2004)], generalizing Sinkhorn's algorithm, shown to be poly time in [Garg, Gurvits, Oliveira, Wigderson 2016] using invariant theory
- Algebraic poly time algorithms [Ivanyos, Qiao, Subrahmanyam 2017] works in any char

- Tensor action: Alternating minimization algorithm by [Verstrate et al 2004] analyzed by [B, Franks, Garg, Oliveira, Walter, Wigderson 2018]. However, only get exponential time algorithm!

Are there poly time algorithms for null cone memberships for tensors?

Null cone membership: previously known results

- Simultaneous LR-action: $G = \mathrm{SL}_n \times \mathrm{SL}_n$ acts on $V = (\mathbb{C}^{n\times n})^m$:

$$\pi(g,h)(A_1,\ldots,A_m) := (gA_1h^T,\ldots,gA_mh^T)$$

- Alternating minimization algorithm by [Gurvits 2004)], generalizing Sinkhorn's algorithm, shown to be poly time in [Garg, Gurvits, Oliveira, Wigderson 2016] using invariant theory
- Algebraic poly time algorithms [Ivanyos, Qiao, Subrahmanyam 2017] works in any char

- Tensor action: Alternating minimization algorithm by [Verstrate et al 2004] analyzed by [B, Franks, Garg, Oliveira, Walter, Wigderson 2018]. However, only get exponential time algorithm!

Are there poly time algorithms for null cone memberships for tensors?

Null cone membership: previously known results

- Simultaneous LR-action: $G = \mathrm{SL}_n \times \mathrm{SL}_n$ acts on $V = (\mathbb{C}^{n\times n})^m$:

$$\pi(g,h)(A_1,\ldots,A_m) := (gA_1h^T,\ldots,gA_mh^T)$$

- Alternating minimization algorithm by [Gurvits 2004)], generalizing Sinkhorn's algorithm, shown to be poly time in [Garg, Gurvits, Oliveira, Wigderson 2016] using invariant theory
- Algebraic poly time algorithms [Ivanyos, Qiao, Subrahmanyam 2017] works in any char

- Tensor action: Alternating minimization algorithm by [Verstrate et al 2004] analyzed by [B, Franks, Garg, Oliveira, Walter, Wigderson 2018]. However, only get exponential time algorithm!

Are there poly time algorithms for null cone memberships for tensors?

Null cone membership: previously known results

- Simultaneous LR-action: $G = \mathrm{SL}_n \times \mathrm{SL}_n$ acts on $V = (\mathbb{C}^{n \times n})^m$:

$$\pi(g,h)(A_1,\ldots,A_m) := (gA_1h^T,\ldots,gA_mh^T)$$

- Alternating minimization algorithm by [Gurvits 2004)], generalizing Sinkhorn's algorithm, shown to be poly time in [Garg, Gurvits, Oliveira, Wigderson 2016] using invariant theory
- Algebraic poly time algorithms [Ivanyos, Qiao, Subrahmanyam 2017] works in any char

- Tensor action: Alternating minimization algorithm by [Verstrate et al 2004] analyzed by [B, Franks, Garg, Oliveira, Walter, Wigderson 2018]. However, only get exponential time algorithm!

Are there poly time algorithms for null cone memberships for tensors?

Second order algorithm

Generalize [Allen-Zhu, Garg, Li, Oliveira, Wigderson 18] from L/R action.

For given $v \in V \setminus 0$ and $\kappa, \varepsilon > 0$, add to F_v a regularizing term $F(g) := F_v(g) + \frac{\varepsilon}{\kappa} \operatorname{reg}(g)$, where $\operatorname{reg}(g) := \|g\|_F^2 + \|g^{-1}\|_F^2$ upper bounds condition number $\kappa_F(g) \leq \operatorname{reg}(g)$.

Algorithm:

- Set $g_0 = I$.
- For $t = 0, \ldots, T-1$:
  1. Compute the geodesic gradient $L := \nabla F(g_t)$ and Hessian $Q := \nabla^2 F(g_t)$ at g_t.
  2. Solve the following (Euclidean) convex quadratic optimization problem:

$$H_t := \operatorname{argmin}\left\{\operatorname{tr}[LH] + \frac{1}{2e}\operatorname{tr}[Q(H \otimes H)] \;:\; H \in i\mathrm{Lie}(K),\ \|H\|_F \leq \frac{1}{4N(\pi)}\right\}$$

  3. Set $g_{t+1} := e^{H_t/e^2} g_t$.
- **Return** g_T.

Second order algorithm

Generalize [Allen-Zhu, Garg, Li, Oliveira, Wigderson 18] from L/R action.

For given $v \in V \setminus 0$ and $\kappa, \varepsilon > 0$, add to F_v a regularizing term $F(g) := F_v(g) + \frac{\varepsilon}{\kappa} \operatorname{reg}(g)$, where $\operatorname{reg}(g) := \|g\|_F^2 + \|g^{-1}\|_F^2$ upper bounds condition number $\kappa_F(g) \le \operatorname{reg}(g)$.

Algorithm:

- Set $g_0 = I$.
- For $t = 0, \ldots, T-1$:
  1. Compute the geodesic gradient $L := \nabla F(g_t)$ and Hessian $Q := \nabla^2 F(g_t)$ at g_t.
  2. Solve the following (Euclidean) convex quadratic optimization problem:

$$H_t := \operatorname{argmin} \left\{ \operatorname{tr}[LH] + \frac{1}{2e} \operatorname{tr}[Q(H \otimes H)] \; : \; H \in i\mathrm{Lie}(K), \; \|H\|_F \le \frac{1}{4N(\pi)} \right\}$$

  3. Set $g_{t+1} := e^{H_t/e^2} g_t$.
- **Return** g_T.

Second order algorithm

Generalize [Allen-Zhu, Garg, Li, Oliveira, Wigderson 18] from L/R action.

For given $v \in V \setminus 0$ and $\kappa, \varepsilon > 0$, add to F_v a regularizing term $F(g) := F_v(g) + \frac{\varepsilon}{\kappa} \operatorname{reg}(g)$, where $\operatorname{reg}(g) := \|g\|_F^2 + \|g^{-1}\|_F^2$ upper bounds condition number $\kappa_F(g) \leq \operatorname{reg}(g)$.

Algorithm:

- Set $g_0 = I$.
- For $t = 0, \ldots, T-1$:
  1. Compute the geodesic gradient $L := \nabla F(g_t)$ and Hessian $Q := \nabla^2 F(g_t)$ at g_t.
  2. Solve the following (Euclidean) convex quadratic optimization problem:

$$H_t := \operatorname{argmin} \left\{ \operatorname{tr}[LH] + \frac{1}{2e} \operatorname{tr}[Q(H \otimes H)] \;:\; H \in i\mathrm{Lie}(K),\ \|H\|_F \leq \frac{1}{4N(\pi)} \right\}$$

  3. Set $g_{t+1} := e^{H_t/e^2} g_t$.
- **Return** g_T.

Analysis of second order algorithm

- Assume $\mathrm{cap}(v) > 0$ and put $C := \log(\|v\|/\mathrm{cap}(v))$.
- Suppose there is a well-conditioned approximate minimizer $g_\star$ s.t.

$$\log\|\pi(g_\star)v\| \leq \log\mathrm{cap}(v) + \varepsilon, \quad \mathrm{reg}(g_\star) \leq \kappa.$$

- Then the second order algorithm, with #iterations

$$T \geq 8e^2 N(\pi)\sqrt{n}\left(\log\kappa + \log\left(1 + \frac{C}{\varepsilon}\right)\right)\log\left(\frac{C}{\varepsilon}\right)$$

returns a group element $g \in G = \mathrm{GL}_n$ such that

$$\log\|\pi(g)v\| \leq \log\mathrm{cap}(v) + 3\varepsilon.$$

- Show existence of well-conditioned approximate minimizer by analyzing the flow of gradient vector field of $G \to \mathbb{R}$, $g \mapsto \frac{1}{2}\|\pi(g)v\|^2$.
- Obtain running time polynomial in $\gamma(\pi)^{-1}$, $\log\varepsilon^{-1}$, while gradient descent had polynomials dependence on ε^{-1}.
- Due to $\gamma(\pi)^{-1}$ dependence still don't get poly time for null cone membership.

Analysis of second order algorithm

- Assume $\mathrm{cap}(v) > 0$ and put $C := \log(\|v\|/\mathrm{cap}(v))$.
- Suppose there is a well-conditioned approximate minimizer $g_\star$ s.t.

$$\log\|\pi(g_\star)v\| \leq \log\mathrm{cap}(v) + \varepsilon, \quad \mathrm{reg}(g_\star) \leq \kappa.$$

- Then the second order algorithm, with #iterations

$$T \geq 8e^2 N(\pi)\sqrt{n}\left(\log\kappa + \log\left(1 + \frac{C}{\varepsilon}\right)\right)\log\left(\frac{C}{\varepsilon}\right)$$

  returns a group element $g \in G = \mathrm{GL}_n$ such that

$$\log\|\pi(g)v\| \leq \log\mathrm{cap}(v) + 3\varepsilon.$$

- Show existence of well-conditioned approximate minimizer by analyzing the flow of gradient vector field of $G \to \mathbb{R}$, $g \mapsto \frac{1}{2}\|\pi(g)v\|^2$.
- Obtain running time polynomial in $\gamma(\pi)^{-1}$, $\log\varepsilon^{-1}$, while gradient descent had polynomials dependence on ε^{-1}.
- Due to $\gamma(\pi)^{-1}$ dependence still don't get poly time for null cone membership.

Analysis of second order algorithm

- Assume $\mathrm{cap}(v) > 0$ and put $C := \log(\|v\|/\mathrm{cap}(v))$.
- Suppose there is a well-conditioned approximate minimizer $g_\star$ s.t.

$$\log\|\pi(g_\star)v\| \leq \log\mathrm{cap}(v) + \varepsilon, \quad \mathrm{reg}(g_\star) \leq \kappa.$$

- Then the second order algorithm, with #iterations

$$T \geq 8e^2 N(\pi)\sqrt{n}\left(\log\kappa + \log\left(1 + \frac{C}{\varepsilon}\right)\right)\log\left(\frac{C}{\varepsilon}\right)$$

  returns a group element $g \in G = \mathrm{GL}_n$ such that

$$\log\|\pi(g)v\| \leq \log\mathrm{cap}(v) + 3\varepsilon.$$

- Show existence of well-conditioned approximate minimizer by analyzing the flow of gradient vector field of $G \to \mathbb{R}$, $g \mapsto \frac{1}{2}\|\pi(g)v\|^2$.
- Obtain running time polynomial in $\gamma(\pi)^{-1}$, $\log\varepsilon^{-1}$, while gradient descent had polynomials dependence on ε^{-1}.
- Due to $\gamma(\pi)^{-1}$ dependence still don't get poly time for null cone membership.

Open problems

- Is the null cone membership problem for general group actions in P? E.g., for tensor action?
- Intermediate goal: in NP $\cap$ coNP?
- Same question may be asked about the moment polytope membership problem for general group actions. Note that membership in "global moment polytopes" is known to be in NP $\cap$ coNP [BCMW17].
- In the non-commutative case, our algorithms' guarantees do not match those of ellipsoid or interior point methods.
- Can we extend non-commutative/geodesic optimization to include interior point methods? (Possible in the commutative case [B, Li, Nieuwboer, Walter 2020].)
- Can geodesic optimization lead to new efficient algorithms in combinatorial optimization?

Thank you for listening