Efficient algorithms for tensor scaling, quantum marginals, and moment polytopes

Cole Franks (RUTGERS)

based on joint work with
Peter Bürgisser, Ankit Garg, Rafael Oliveira, Michael Walter, Avi Wigderson

Overview

- Simple classical algorithm for tensor scaling
- Important example of moment polytope problem
- Analysis solves nonconvex optimization problem arising in GIT
- Many interesting consequences of faster algorithms

Overview

- Simple classical algorithm for tensor scaling
- Important example of moment polytope problem
- Analysis solves nonconvex optimization problem arising in GIT
- Many interesting consequences of faster algorithms

Overview

- Simple classical algorithm for tensor scaling
- Important example of moment polytope problem
- Analysis solves nonconvex optimization problem arising in GIT
- Many interesting consequences of faster algorithms

Overview

- Simple classical algorithm for tensor scaling
- Important example of moment polytope problem
- Analysis solves nonconvex optimization problem arising in GIT
- Many interesting consequences of faster algorithms

Outline

- Problem statement and history
- Algorithm
- Analysis
- Conclusion and open problems
- More moment polytopes

Outline

- Problem statement and history
- Algorithm
- Analysis
- Conclusion and open problems
- More moment polytopes

Outline

- Problem statement and history
- Algorithm
- Analysis
- Conclusion and open problems
- More moment polytopes

Outline

- Problem statement and history
- Algorithm
- Analysis
- Conclusion and open problems
- More moment polytopes

Problem statement and history

Quantum marginal problems

Space of d-tensors, denoted $\mathbb{C}^{n_1} \otimes \mathbb{C}^{n_2} \otimes \cdots \otimes \mathbb{C}^{n_d}$:

d-dimensional complex arrays of dimensions $n_1, \ldots, n_d$; entries

$$x_{i_1,\ldots,i_d} \in \mathbb{C}$$

for $i_j \in [n_j]$. Let $n = n_1 \ldots n_d$.

e.g. $n_1 = n_2 = n_3 = 2$:

$X =$

Quantum marginal problems

Space of d-tensors, denoted $\mathbb{C}^{n_1} \otimes \mathbb{C}^{n_2} \otimes \cdots \otimes \mathbb{C}^{n_d}$:

d-dimensional complex arrays of dimensions $n_1, \ldots, n_d$; entries

$$x_{i_1,\ldots,i_d} \in \mathbb{C}$$

for $i_j \in [n_j]$. Let $n = n_1 \ldots n_d$.

e.g. $n_1 = n_2 = n_3 = 2$:

$X =$

Quantum marginal problems

Space of d-tensors, denoted $\mathbb{C}^{n_1} \otimes \mathbb{C}^{n_2} \otimes \cdots \otimes \mathbb{C}^{n_d}$:

d-dimensional complex arrays of dimensions $n_1, \ldots, n_d$; entries

$$x_{i_1,\ldots,i_d} \in \mathbb{C}$$

for $i_j \in [n_j]$. Let $n = n_1 \ldots n_d$.

e.g. $n_1 = n_2 = n_3 = 2$:

$X =$

Quantum marginal problems

Space of d-tensors, denoted $\mathbb{C}^{n_1} \otimes \mathbb{C}^{n_2} \otimes \cdots \otimes \mathbb{C}^{n_d}$:

d-dimensional complex arrays of dimensions $n_1, \ldots, n_d$; entries

$$x_{i_1,\ldots,i_d} \in \mathbb{C}$$

for $i_j \in [n_j]$. Let $n = n_1 \ldots n_d$.

e.g. $n_1 = n_2 = n_3 = 2$:

$X =$

Quantum marginal problems

Space of d-tensors, denoted $\mathbb{C}^{n_1} \otimes \mathbb{C}^{n_2} \otimes \cdots \otimes \mathbb{C}^{n_d}$:

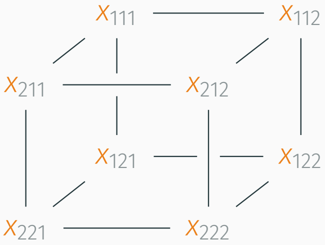
d-dimensional complex arrays of dimensions $n_1, \ldots, n_d$; entries

$$x_{i_1,\ldots,i_d} \in \mathbb{C}$$

for $i_j \in [n_j]$. Let $n = n_1 \ldots n_d$.

e.g. $n_1 = n_2 = n_3 = 2$:

$X =$



Marginals of a tensor

- Let X be a d-tensor
- Consider j^{th} slice in i^{th} direction:

$$X_{\underbrace{**\cdots**}_{i-1}j\underbrace{**\cdots**}_{d-i}}$$

  it is a $(d-1)$-tensor.
- The i^{th} marginal $\rho_X^{(i)}$ is the $n_i \times n_i$ Gram matrix of the slices in the i^{th} direction.

Marginals of a tensor

- Let X be a d-tensor
- Consider j^{th} slice in i^{th} direction:

$$X_{\underbrace{**\cdots**}_{i-1}j\underbrace{**\cdots**}_{d-i}}$$

  it is a $(d-1)$-tensor.

- The i^{th} marginal $\rho_X^{(i)}$ is the $n_i \times n_i$ Gram matrix of the slices in the i^{th} direction.

Marginals of a tensor

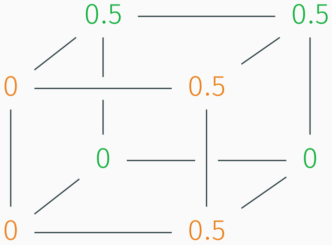
- Let X be a d-tensor
- Consider j^{th} slice in i^{th} direction:

$$X_{\underbrace{**\cdots**}_{i-1}j\underbrace{**\cdots**}_{d-i}}$$

  it is a $(d-1)$-tensor.
- The i^{th} marginal $\rho_X^{(i)}$ is the $n_i \times n_i$ Gram matrix of the slices in the i^{th} direction.

Example: $n_1 = n_2 = n_3 = 2$

$X =$



$$X_{1**} = \begin{bmatrix} 0.5 & 0.5 \\ 0 & 0 \end{bmatrix}$$

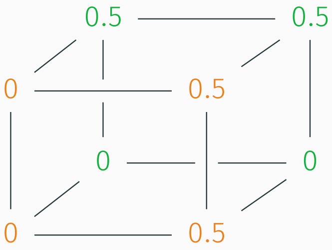
$$X_{2**} = \begin{bmatrix} 0 & 0.5 \\ 0 & 0.5 \end{bmatrix}$$

$$\rho_X^{(1)} := \begin{bmatrix} - X_{1**} - \\ - X_{2**} - \end{bmatrix} \left(\begin{bmatrix} - X_{1**} - \\ - X_{2**} - \end{bmatrix}^\dagger \right) = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.5 \end{bmatrix}$$

Note: $\operatorname{Tr} \rho_X^{(i)} = \|X\|^2$!

Example: $n_1 = n_2 = n_3 = 2$

$X =$



$$X_{1**} = \begin{bmatrix} 0.5 & 0.5 \\ 0 & 0 \end{bmatrix}$$

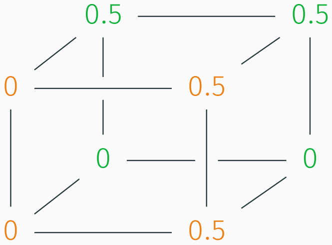
$$X_{2**} = \begin{bmatrix} 0 & 0.5 \\ 0 & 0.5 \end{bmatrix}$$

$$\rho_X^{(1)} := \begin{bmatrix} - X_{1**} - \\ - X_{2**} - \end{bmatrix} \left(\begin{bmatrix} - X_{1**} - \\ - X_{2**} - \end{bmatrix}^\dagger \right) = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.5 \end{bmatrix}$$

Note: $\mathbf{Tr}\, \rho_X^{(i)} = \|X\|^2$!

Example: $n_1 = n_2 = n_3 = 2$

$X =$



$$X_{1**} = \begin{bmatrix} 0.5 & 0.5 \\ 0 & 0 \end{bmatrix}$$

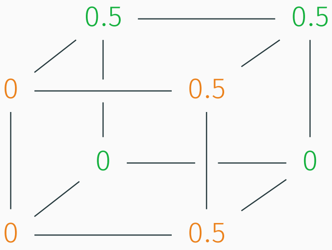
$$X_{2**} = \begin{bmatrix} 0 & 0.5 \\ 0 & 0.5 \end{bmatrix}$$

$$\rho_X^{(1)} := \begin{bmatrix} - X_{1**} - \\ - X_{2**} - \end{bmatrix} \left(\begin{bmatrix} - X_{1**} - \\ - X_{2**} - \end{bmatrix}^{\dagger} \right) = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.5 \end{bmatrix}$$

Note: $\operatorname{Tr} \rho_X^{(i)} = \|X\|^2$!

Example: $n_1 = n_2 = n_3 = 2$

$X =$



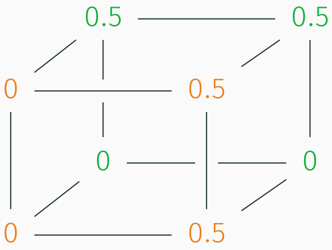
$$X_{1**} = \begin{bmatrix} 0.5 & 0.5 \\ 0 & 0 \end{bmatrix}$$

$$X_{2**} = \begin{bmatrix} 0 & 0.5 \\ 0 & 0.5 \end{bmatrix}$$

$$\rho_X^{(1)} := \begin{bmatrix} - X_{1**} - \\ - X_{2**} - \end{bmatrix} \left(\begin{bmatrix} - X_{1**} - \\ - X_{2**} - \end{bmatrix}^{\dagger} \right) = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.5 \end{bmatrix}$$

Note: $\operatorname{Tr} \rho_X^{(i)} = \|X\|^2$!

$X =$



$$X_{1**} = \begin{bmatrix} 0.5 & 0.5 \\ 0 & 0 \end{bmatrix}$$

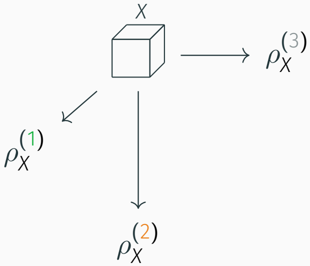
$$X_{2**} = \begin{bmatrix} 0 & 0.5 \\ 0 & 0.5 \end{bmatrix}$$

$$\rho_X^{(1)} := \begin{bmatrix} - & X_{1**} & - \\ - & X_{2**} & - \end{bmatrix} \left(\begin{bmatrix} - & X_{1**} & - \\ - & X_{2**} & - \end{bmatrix}^{\dagger} \right) = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.5 \end{bmatrix}$$

Note: $\operatorname{Tr} \rho_X^{(i)} = \|X\|^2$!

Interpretation

If Alice, Bob, and Carol each hold a qubit but the joint state is X, $\rho_X^{(1)}, \rho_X^{(2)}, \rho_X^{(3)}$ are the mixed states of their respective qubits.



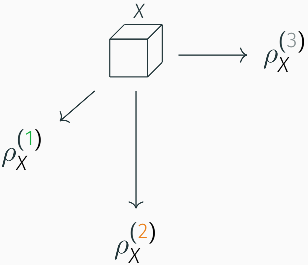
One body quantum marginal problem, $d = 3$:

Can PSD matrices A, B, C arise as the marginals of some tensor X?

Fact: the answer depends only on $\mathrm{spec}(A), \mathrm{spec}(B), \mathrm{spec}(C)$.

Interpretation

If Alice, Bob, and Carol each hold a qubit but the joint state is X, $\rho_X^{(1)}, \rho_X^{(2)}, \rho_X^{(3)}$ are the mixed states of their respective qubits.



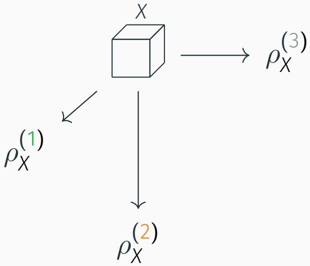
One body quantum marginal problem, $d = 3$:

Can PSD matrices A, B, C arise as the marginals of some tensor X?

Fact: the answer depends only on $\mathrm{spec}(A), \mathrm{spec}(B), \mathrm{spec}(C)$.

Interpretation

If Alice, Bob, and Carol each hold a qubit but the joint state is X, $\rho_X^{(1)}, \rho_X^{(2)}, \rho_X^{(3)}$ are the mixed states of their respective qubits.



One body quantum marginal problem, $d = 3$:

Can PSD matrices A, B, C arise as the marginals of some tensor X?

Fact: the answer depends only on $\mathrm{spec}(A), \mathrm{spec}(B), \mathrm{spec}(C)$.

Convenient notation

If $\mathbb{X} \subset \mathbb{C}^{n_0} \otimes \cdots \otimes \mathbb{C}^{n_d}$ is a set of $d+1$-tensors, let

$$\Delta(\mathbb{X}) = \left\{ \left(\operatorname{spec}(\rho_Y^{(1)})/\|Y\|^2, \ldots, \operatorname{spec}(\rho_Y^{(d)})/\|Y\|^2 \right) : Y \in \mathbb{X} \right\}$$

$\Delta(\mathbb{X})$ *is all the tuples of spectra of marginals of elements of* $\mathbb{X}$*, normalized to have trace one.*

Quantum marginal problem, restatement:

Input: $p = (p_1, p_2, p_3)$ list of sequences of nonnegative reals

Output: Whether $p \in \Delta(\mathbb{C}^{n_0=1} \otimes \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d})$.

Convenient notation

If $\mathbb{X} \subset \mathbb{C}^{n_0} \otimes \cdots \otimes \mathbb{C}^{n_d}$ is a set of $d+1$-tensors, let

$$\Delta(\mathbb{X}) = \left\{ \left(\mathrm{spec}(\rho_Y^{(1)})/\|Y\|^2, \ldots, \mathrm{spec}(\rho_Y^{(d)})/\|Y\|^2 \right) : Y \in \mathbb{X} \right\}$$

$\Delta(\mathbb{X})$ *is all the tuples of spectra of marginals of elements of* $\mathbb{X}$*, normalized to have trace one.*

Quantum marginal problem, restatement:

Input: $p = (p_1, p_2, p_3)$ list of sequences of nonnegative reals

Output: Whether $p \in \Delta(\mathbb{C}^{n_0=1} \otimes \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d})$.

Convenient notation

If $\mathbb{X} \subset \mathbb{C}^{n_0} \otimes \cdots \otimes \mathbb{C}^{n_d}$ is a set of $d+1$-tensors, let

$$\Delta(\mathbb{X}) = \left\{ \left(\mathrm{spec}(\rho_Y^{(1)})/\|Y\|^2, \ldots, \mathrm{spec}(\rho_Y^{(d)})/\|Y\|^2 \right) : Y \in \mathbb{X} \right\}$$

$\Delta(\mathbb{X})$ *is all the tuples of spectra of marginals of elements of* $\mathbb{X}$*, normalized to have trace one.*

Quantum marginal problem, restatement:

Input: $p = (p_1, p_2, p_3)$ list of sequences of nonnegative reals

Output: Whether $p \in \Delta(\mathbb{C}^{n_0=1} \otimes \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d})$.

More generally:

Given a tensor X, can we locally change basis to obtain specific marginals?

We consider a $d+1$ tensor $X \in \mathbb{C}^{n_0} \otimes \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$, and let $G := \mathrm{GL}_{n_1} \times \cdots \times \mathrm{GL}_{n_d}$.

$$g \cdot X := (I_{n_0} \otimes g_1 \otimes g_2 \otimes \cdots \otimes g_d) X.$$

$G \cdot X$ denotes the *orbit* of X, and $\overline{G \cdot X}$ the *orbit closure*.

Question: TENSORSCALING(X, p)

Input: $p = (p_1, \ldots, p_d)$,
X a tensor in $\mathbb{C}^{n_0} \otimes \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$

Output: whether $p \in \Delta(\overline{G \cdot X})$.

More generally:

Given a tensor X, can we locally change basis to obtain specific marginals?

We consider a $d+1$ tensor $X \in \mathbb{C}^{n_0} \otimes \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$, and let $G := \mathrm{GL}_{n_1} \times \cdots \times \mathrm{GL}_{n_d}$.

$$g \cdot X := (I_{n_0} \otimes g_1 \otimes g_2 \otimes \cdots \otimes g_d)X.$$

$G \cdot X$ denotes the *orbit* of X, and $\overline{G \cdot X}$ the *orbit closure*.

Question: TENSORSCALING(X, p)

Input: $p = (p_1, \ldots, p_d)$,
X a tensor in $\mathbb{C}^{n_0} \otimes \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$

Output: whether $p \in \Delta(\overline{G \cdot X})$.

More generally:

Given a tensor X, can we locally change basis to obtain specific marginals?

We consider a $d+1$ tensor $X \in \mathbb{C}^{n_0} \otimes \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$, and let $G := \mathrm{GL}_{n_1} \times \cdots \times \mathrm{GL}_{n_d}$.

$$g \cdot X := (I_{n_0} \otimes g_1 \otimes g_2 \otimes \cdots \otimes g_d)X.$$

$G \cdot X$ denotes the *orbit* of X, and $\overline{G \cdot X}$ the *orbit closure*.

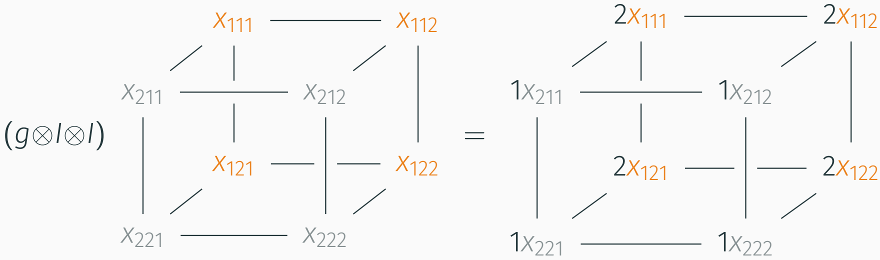
Question: TENSORSCALING(X, p)

Input: $p = (p_1, \ldots, p_d)$,
X a tensor in $\mathbb{C}^{n_0} \otimes \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$

Output: whether $p \in \Delta(\overline{G \cdot X})$.

Example

E.g. if $g = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ then



Moment polytopes

Amazing fact:

$\Delta(\mathbb{C}^{n_0} \otimes \cdots \otimes \mathbb{C}^{n_d})$ and $\Delta(\overline{G \cdot X})$ are convex polytopes!

More generally: Holds if $\mathbb{X}$ is a variety and $G \cdot \mathbb{X} \subset \mathbb{X}$. Then $\Delta(\mathbb{X})$ is called the *moment polytope* for the action of G on $\mathbb{X}$.

The groups can also be more general.

Moment polytopes

Amazing fact:

$\Delta(\mathbb{C}^{n_0} \otimes \cdots \otimes \mathbb{C}^{n_d})$ and $\Delta(\overline{G \cdot X})$ are convex polytopes!

More generally: Holds if $\mathbb{X}$ is a variety and $G \cdot \mathbb{X} \subset \mathbb{X}$. Then $\Delta(\mathbb{X})$ is called the *moment polytope* for the action of G on $\mathbb{X}$.

The groups can also be more general.

Moment polytopes

Amazing fact:

$\Delta(\mathbb{C}^{n_0} \otimes \cdots \otimes \mathbb{C}^{n_d})$ and $\Delta(\overline{G \cdot X})$ are convex polytopes!

More generally: Holds if $\mathbb{X}$ is a variety and $G \cdot \mathbb{X} \subset \mathbb{X}$. Then $\Delta(\mathbb{X})$ is called the *moment polytope* for the action of G on $\mathbb{X}$.

The groups can also be more general.

Example: reducing matrix scaling to tensor scaling

Question: MATRIXSCALING(A)

Input: A, nonnegative matrix

Output: Whether $\exists D_1, D_2 \succ 0$ diagonal with $D_1 A D_2$ *doubly stochastic.*

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a nonnegative matrix, and let

$p_1 = p_2 = (1, 1)$ and $X =$

Fact: MATRIXSCALING(A) $\iff$ TENSORSCALING(X, p)

Example: reducing matrix scaling to tensor scaling

Question: MATRIXSCALING(A)

Input: A, nonnegative matrix

Output: Whether $\exists D_1, D_2 \succ 0$ diagonal with $D_1 A D_2$ *doubly stochastic.*

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a nonnegative matrix, and let

$p_1 = p_2 = (1,1)$ and $X =$

Fact: MATRIXSCALING(A) $\iff$ TENSORSCALING(X, p)

Example: reducing matrix scaling to tensor scaling

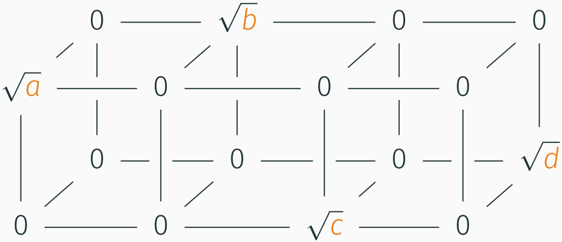
Question: MATRIXSCALING(A)

Input: A, nonnegative matrix

Output: Whether $\exists D_1, D_2 \succ 0$ diagonal with $D_1 A D_2$ *doubly stochastic.*

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a nonnegative matrix, and let

$\boldsymbol{p}_1 = \boldsymbol{p}_2 = (1,1)$ and $X =$



Fact: MATRIXSCALING(A) $\iff$ TENSORSCALING($X, \boldsymbol{p}$)

Example: reducing matrix scaling to tensor scaling

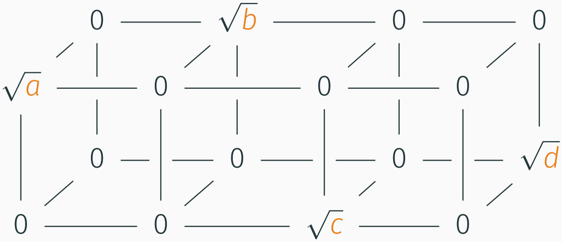
Question: MATRIXSCALING(A)

Input: A, nonnegative matrix

Output: Whether $\exists D_1, D_2 \succ 0$ diagonal with $D_1 A D_2$ *doubly stochastic.*

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a nonnegative matrix, and let

$p_1 = p_2 = (1, 1)$ and $X =$



Fact: MATRIXSCALING(A) $\iff$ TENSORSCALING(X, p)

Applications of tensor scaling

Matrix scaling:

- deterministically approximating permanent

Operator scaling: (The $d = 2$ case of tensor scaling)

- noncommutative rational identity testing
- Forster's radial isotropic position
- computing the Brascamp-Lieb constant in analysis
- Horn's problem on eigenvalues of sums of matrices

One body quantum marginal problem: (Tensor scaling for random X)

- equivalence under SLOCC to locally maximally mixed state
- The *Kronecker polytope* in representation theory

Tensor scaling:

- null-cone: do all $\mathrm{SL}_{n_1} \times \ldots \mathrm{SL}_{n_d}$-invariant polynomials vanish on X?

Applications of tensor scaling

Matrix scaling:

- deterministically approximating permanent

Operator scaling: (The $d = 2$ case of tensor scaling)

- noncommutative rational identity testing
- Forster's radial isotropic position
- computing the Brascamp-Lieb constant in analysis
- Horn's problem on eigenvalues of sums of matrices

One body quantum marginal problem: (Tensor scaling for random X)

- equivalence under SLOCC to locally maximally mixed state
- The *Kronecker polytope* in representation theory

Tensor scaling:

- null-cone: do all $\mathrm{SL}_{n_1} \times \ldots \mathrm{SL}_{n_d}$-invariant polynomials vanish on X?

Applications of tensor scaling

Matrix scaling:

- deterministically approximating permanent

Operator scaling: (The $d = 2$ case of tensor scaling)

- noncommutative rational identity testing
- Forster's radial isotropic position
- computing the Brascamp-Lieb constant in analysis
- Horn's problem on eigenvalues of sums of matrices

One body quantum marginal problem: (Tensor scaling for random X)

- equivalence under SLOCC to locally maximally mixed state
- The *Kronecker polytope* in representation theory

Tensor scaling:

- null-cone: do all $\mathrm{SL}_{n_1} \times \ldots \mathrm{SL}_{n_d}$-invariant polynomials vanish on X?

Applications of tensor scaling

Matrix scaling:

- deterministically approximating permanent

Operator scaling: (The $d = 2$ case of tensor scaling)

- noncommutative rational identity testing
- Forster's radial isotropic position
- computing the Brascamp-Lieb constant in analysis
- Horn's problem on eigenvalues of sums of matrices

One body quantum marginal problem: (Tensor scaling for random X)

- equivalence under SLOCC to locally maximally mixed state
- The *Kronecker polytope* in representation theory

Tensor scaling:

- null-cone: do all $\mathrm{SL}_{n_1} \times \ldots \mathrm{SL}_{n_d}$-invariant polynomials vanish on X?

Approximate tensor scaling: $\text{TENSORSCALING}(X, \boldsymbol{p}, \varepsilon)$

Input: Tensor X, tuple $\boldsymbol{p}$, $\varepsilon > 0$

Output: If either output g such that for all $i \in [d]$

$$\| \operatorname{spec}(\rho^{(i)}_{g \cdot X}) - \boldsymbol{p}_i \|_1 \leq \varepsilon,$$

or correctly output that $\boldsymbol{p} \notin \Delta(\overline{G \cdot X})$.

History of approxmate scaling algorithms

MATRIXSCALING(A, r, c):

- [Sinkhorn '64]: simple $\mathrm{poly}(1/\varepsilon)$ algorithm when $r = c = 1$
- [Linial, Samorodnitsky, Wigderson '98]: $\mathrm{poly}\log(1/\varepsilon)$ for any r, c

OPERATORSCALING(X, p_1, p_2): The $d = 2$ case of TENSORSCALING

- [Gurvits '04]: simple $\mathrm{poly}(1/\varepsilon)$ algorithm when $p_1 = p_2 = 1$
- [GGOW'17]: decision problem $p_1 = p_2 = 1$
- [AGLOW'18]: $\mathrm{poly}\log(1/\varepsilon)$ for $p_1 = p_2 = 1$
- [_'18]: randomized $\mathrm{poly}(1/\varepsilon)$ for any p_1, p_2

One body quantum marginal problem:

- [BCMW'17]: decision problem is in $\mathrm{NP} \cap \mathrm{coNP}$

TENSORSCALING($X, p_1, \ldots, p_d$):

- [BGOWW'17]: $\mathrm{poly}(1/\varepsilon)$ for $p_i = 1/n_i$
- [BFGOWW'17]: (this work:) randomized $\mathrm{poly}(1/\varepsilon)$ for any $p_1, \ldots, p_d$

History of approxmate scaling algorithms

MATRIXSCALING(A, r, c):

- [Sinkhorn '64]: simple $\mathrm{poly}(1/\varepsilon)$ algorithm when $r = c = 1$
- [Linial, Samorodnitsky, Wigderson '98]: $\mathrm{poly}\log(1/\varepsilon)$ for any r, c

OPERATORSCALING(X, p_1, p_2): The $d = 2$ case of TENSORSCALING

- [Gurvits '04]: simple $\mathrm{poly}(1/\varepsilon)$ algorithm when $p_1 = p_2 = 1$
- [GGOW'17]: decision problem $p_1 = p_2 = 1$
- [AGLOW'18]: $\mathrm{poly}\log(1/\varepsilon)$ for $p_1 = p_2 = 1$
- [_'18]: randomized $\mathrm{poly}(1/\varepsilon)$ for any p_1, p_2

One body quantum marginal problem:

- [BCMW'17]: decision problem is in $\mathrm{NP} \cap \mathrm{coNP}$

TENSORSCALING($X, p_1, \ldots, p_d$):

- [BGOWW'17]: $\mathrm{poly}(1/\varepsilon)$ for $p_i = 1/n_i$
- [BFGOWW'17]: (this work:) randomized $\mathrm{poly}(1/\varepsilon)$ for any $p_1, \ldots, p_d$

History of approxmate scaling algorithms

MATRIXSCALING(A, r, c):

- [Sinkhorn '64]: simple $\mathrm{poly}(1/\varepsilon)$ algorithm when $r = c = 1$
- [Linial, Samorodnitsky, Wigderson '98]: $\mathrm{poly}\log(1/\varepsilon)$ for any r, c

OPERATORSCALING(X, p_1, p_2): The $d = 2$ case of TENSORSCALING

- [Gurvits '04]: simple $\mathrm{poly}(1/\varepsilon)$ algorithm when $p_1 = p_2 = 1$
- [GGOW'17]: decision problem $p_1 = p_2 = 1$
- [AGLOW'18]: $\mathrm{poly}\log(1/\varepsilon)$ for $p_1 = p_2 = 1$
- [_'18]: randomized $\mathrm{poly}(1/\varepsilon)$ for any p_1, p_2

One body quantum marginal problem:

- [BCMW'17]: decision problem is in $\mathbf{NP} \cap \mathbf{coNP}$

TENSORSCALING$(X, p_1, \ldots, p_d)$:

- [BGOWW'17]: $\mathrm{poly}(1/\varepsilon)$ for $p_i = 1/n_i$
- [BFGOWW'17]: (this work:) randomized $\mathrm{poly}(1/\varepsilon)$ for any $p_1, \ldots, p_d$

History of approxmate scaling algorithms

MATRIXSCALING(A, r, c):

- [Sinkhorn '64]: simple $\text{poly}(1/\varepsilon)$ algorithm when $r = c = 1$
- [Linial, Samorodnitsky, Wigderson '98]: $\text{poly}\log(1/\varepsilon)$ for any r, c

OPERATORSCALING(X, p_1, p_2): The $d = 2$ case of TENSORSCALING

- [Gurvits '04]: simple $\text{poly}(1/\varepsilon)$ algorithm when $p_1 = p_2 = 1$
- [GGOW'17]: decision problem $p_1 = p_2 = 1$
- [AGLOW'18]: $\text{poly}\log(1/\varepsilon)$ for $p_1 = p_2 = 1$
- [_'18]: randomized $\text{poly}(1/\varepsilon)$ for any p_1, p_2

One body quantum marginal problem:

- [BCMW'17]: decision problem is in $\text{NP} \cap \text{coNP}$

TENSORSCALING$(X, p_1, \ldots, p_d)$:

- [BGOWW'17]: $\text{poly}(1/\varepsilon)$ for $p_i = 1/n_i$
- [BFGOWW'17]: (this work:) randomized $\text{poly}(1/\varepsilon)$ for any $p_1, \ldots, p_d$

Our result

Theorem (BFGOWW '18)

There is a randomized $\mathrm{poly}(\langle X\rangle + \langle \mathbf{p}\rangle, 1/\varepsilon)$*-time algorithm for TENSORSCALING(*$X, \mathbf{p}, \varepsilon$*) with success probability* $1/2$.

The algorithm requires

$$O\left(dn^2 \frac{\langle X\rangle + \langle \lambda\rangle + \log dn}{\varepsilon}\right)$$

iterations, each dominated by computing a Cholesky decomposition of some $n_i \times n_i$ matrix.

Implications for decision problem

Convention: $\boldsymbol{p} = \boldsymbol{\lambda}/k$ for $\boldsymbol{\lambda}$ integral and $k = \sum \boldsymbol{\lambda}_j^{(1)}$

Theorem (BFGOWW '18)

If for all i,

$$\| \operatorname{spec}(\rho_{g \cdot X}^{(i)}) - \boldsymbol{p}^{(i)} \|_1 \leq \exp(-O(n_1 + \cdots + n_d) \log k \max n_i),$$

then $\boldsymbol{p} \in \Delta(G \cdot X)$.

Unfortunately, doesn't result in poly time algorithm! Need poly(log(1/ε)).

Implications for decision problem

Convention: $\boldsymbol{p} = \boldsymbol{\lambda}/k$ for $\boldsymbol{\lambda}$ integral and $k = \sum \boldsymbol{\lambda}_j^{(1)}$

Theorem (BFGOWW '18)

If for all i,

$$\| \operatorname{spec}(\rho_{g \cdot X}^{(i)}) - \boldsymbol{p}^{(i)} \|_1 \leq \exp(-O(n_1 + \cdots + n_d) \log k \max n_i),$$

then $\boldsymbol{p} \in \Delta(G \cdot X)$.

Unfortunately, doesn't result in poly time algorithm! Need $\operatorname{poly}(\log(1/\varepsilon))$.

Algorithm

Vague algorithm

Algorithm

Input: $X, \mathbf{p}$ with integer coordinates, ε.

Output: $Y = g \cdot X$ s.t. $\| \mathrm{spec}(\rho_Y^{(i)}) - \mathbf{p}^{(i)} \|_1 \leq \varepsilon$, or OUTSIDE POLYTOPE

- Choose g_0 with i.i.d integer coordinates in $[K]$, set $Y = g_0 \cdot X / \|g_0 \cdot X\|$.
- Repeat T times:
    - If done, output Y.
    - Else, scale in single factor to FIX the worst marginal of Y. (ignoring damage done to other marginals!)
- Output OUTSIDE POLYTOPE

Theorem

For $T \geq \mathrm{poly}(\langle X \rangle + \langle \mathbf{p} \rangle, n, 1/\varepsilon)$, this is algorithm succeeds with probability at least 1/2.

Vague algorithm

Algorithm

Input: $X, \mathbf{p}$ with integer coordinates, ε.

Output: $Y = g \cdot X$ s.t. $\|\mathrm{spec}(\rho_Y^{(i)}) - \mathbf{p}^{(i)}\|_1 \leq \varepsilon$, or OUTSIDE POLYTOPE

- Choose g_0 with i.i.d integer coordinates in $[K]$, set $Y = g_0 \cdot X / \|g_0 \cdot X\|$.
- Repeat T times:
  - If done, output Y.
  - Else, scale in single factor to FIX the worst marginal of Y. (ignoring damage done to other marginals!)
- Output OUTSIDE POLYTOPE

Theorem

For $T \geq \mathrm{poly}(\langle X \rangle + \langle \mathbf{p} \rangle, n, 1/\varepsilon)$, this is algorithm succeeds with probability at least 1/2.

Vague algorithm

Algorithm

Input: $X, \mathbf{p}$ with integer coordinates, ε.

Output: $Y = g \cdot X$ s.t. $\| \mathrm{spec}(\rho_Y^{(i)}) - \mathbf{p}^{(i)} \|_1 \leq \varepsilon$, or OUTSIDE POLYTOPE

- Choose g_0 with i.i.d integer coordinates in $[K]$, set $Y = g_0 \cdot X / \|g_0 \cdot X\|$.
- Repeat T times:
  - If done, output Y.
  - Else, scale in single factor to FIX the worst marginal of Y. (ignoring damage done to other marginals!)
- Output OUTSIDE POLYTOPE

Theorem

For $T \geq \mathrm{poly}(\langle X \rangle + \langle \mathbf{p} \rangle, n, 1/\varepsilon)$, this is algorithm succeeds with probability at least 1/2.

Vague algorithm

Algorithm

Input: $X, \mathbf{p}$ with integer coordinates, ε.

Output: $Y = g \cdot X$ s.t. $\| \mathbf{spec}(\rho_Y^{(i)}) - \mathbf{p}^{(i)} \|_1 \leq \varepsilon$, or OUTSIDE POLYTOPE

- Choose g_0 with i.i.d integer coordinates in $[K]$, set $Y = g_0 \cdot X / \|g_0 \cdot X\|$.
- Repeat T times:
  - If done, output Y.
  - Else, scale in single factor to FIX the worst marginal of Y. (ignoring damage done to other marginals!)
- Output OUTSIDE POLYTOPE

Theorem

For $T \geq \mathbf{poly}(\langle X \rangle + \langle \mathbf{p} \rangle, n, 1/\varepsilon)$, this is algorithm succeeds with probability at least 1/2.

Vague algorithm

Algorithm

Input: $X, \mathbf{p}$ with integer coordinates, ε.

Output: $Y = g \cdot X$ s.t. $\|\mathrm{spec}(\rho_Y^{(i)}) - \mathbf{p}^{(i)}\|_1 \leq \varepsilon$, or OUTSIDE POLYTOPE

- Choose g_0 with i.i.d integer coordinates in $[K]$, set $Y = g_0 \cdot X / \|g_0 \cdot X\|$.
- Repeat T times:
  - If done, output Y.
  - Else, scale in single factor to FIX the worst marginal of Y. (ignoring damage done to other marginals!)
- Output OUTSIDE POLYTOPE

Theorem

For $T \geq \mathrm{poly}(\langle X \rangle + \langle \mathbf{p} \rangle, n, 1/\varepsilon)$, this is algorithm succeeds with probability at least 1/2.

Vague algorithm

Algorithm

Input: $X, \mathbf{p}$ with integer coordinates, ε.

Output: $Y = g \cdot X$ s.t. $\|\mathbf{spec}(\rho_Y^{(i)}) - \mathbf{p}^{(i)}\|_1 \leq \varepsilon$, or OUTSIDE POLYTOPE

- Choose g_0 with i.i.d integer coordinates in $[K]$, set $Y = g_0 \cdot X / \|g_0 \cdot X\|$.
- Repeat T times:
  - If done, output Y.
  - Else, scale in single factor to FIX the worst marginal of Y. (ignoring damage done to other marginals!)
- Output OUTSIDE POLYTOPE

Theorem

For $T \geq \mathbf{poly}(\langle X \rangle + \langle \mathbf{p} \rangle, n, 1/\varepsilon)$, this is algorithm succeeds with probability at least 1/2.

Vague algorithm

Algorithm

Input: $X, \boldsymbol{p}$ with integer coordinates, ε.

Output: $Y = g \cdot X$ s.t. $\| \mathrm{spec}(\rho_Y^{(i)}) - \boldsymbol{p}^{(i)} \|_1 \leq \varepsilon$, or OUTSIDE POLYTOPE

- Choose g_0 with i.i.d integer coordinates in $[K]$, set $Y = g_0 \cdot X / \|g_0 \cdot X\|$.
- Repeat T times:
  - If done, output Y.
  - Else, scale in single factor to FIX the worst marginal of Y. (ignoring damage done to other marginals!)
- Output OUTSIDE POLYTOPE

Theorem

For $T \geq \mathrm{poly}(\langle X \rangle + \langle \boldsymbol{p} \rangle, n, 1/\varepsilon)$, this is algorithm succeeds with probability at least $1/2$.

Scaling II

$g = (g_1, I, \ldots, I)$ scales the flattening:

$$\begin{bmatrix} - (g \cdot Y)_{1*\cdots} - \\ \vdots \\ - (g \cdot Y)_{n_1*\cdots} - \end{bmatrix} = g_1 \begin{bmatrix} - Y_{1*\cdots} - \\ \vdots \\ - X_{n_1*\cdots} - \end{bmatrix}$$

In particular,

$$\rho^{(1)}_{g\cdot Y} = g_1 \begin{bmatrix} - Y_{1*\cdots} - \\ \vdots \\ - X_{n_1*\cdots} - \end{bmatrix} \left(g_1 \begin{bmatrix} - Y_{1*\cdots} - \\ \vdots \\ - X_{n_1*\cdots} - \end{bmatrix} \right)^\dagger = g_1 \rho^{(1)}_Y g_1^\dagger$$

Scaling II

$g = (g_1, I, \ldots, I)$ scales the flattening:

$$\begin{bmatrix} - (g \cdot Y)_{1*\cdots} - \\ \vdots \\ - (g \cdot Y)_{n_1*\cdots} - \end{bmatrix} = g_1 \begin{bmatrix} - Y_{1*\cdots} - \\ \vdots \\ - X_{n_1*\cdots} - \end{bmatrix}$$

In particular,

$$\rho^{(1)}_{g\cdot Y} = g_1 \begin{bmatrix} - Y_{1*\cdots} - \\ \vdots \\ - X_{n_1*\cdots} - \end{bmatrix} \left(g_1 \begin{bmatrix} - Y_{1*\cdots} - \\ \vdots \\ - X_{n_1*\cdots} - \end{bmatrix} \right)^{\dagger} = g_1 \rho^{(1)}_Y g_1^{\dagger}$$

Scaling II

$g = (g_1, I, \ldots, I)$ scales the flattening:

$$\begin{bmatrix} - (g \cdot Y)_{1*\cdots} - \\ \vdots \\ - (g \cdot Y)_{n_1*\cdots} - \end{bmatrix} = g_1 \begin{bmatrix} - Y_{1*\cdots} - \\ \vdots \\ - X_{n_1*\cdots} - \end{bmatrix}$$

In particular,

$$\rho^{(1)}_{g \cdot Y} = g_1 \begin{bmatrix} - Y_{1*\cdots} - \\ \vdots \\ - X_{n_1*\cdots} - \end{bmatrix} \left(g_1 \begin{bmatrix} - Y_{1*\cdots} - \\ \vdots \\ - X_{n_1*\cdots} - \end{bmatrix} \right)^{\dagger} = g_1 \rho^{(1)}_Y g_1^{\dagger}$$

Fixing a marginal

Easy to fix i^{th} marginal: choose g_i such that $g_i \rho_Y^{(i)} g_i^\dagger = \mathrm{diag}(p^{(i)})$.

WARNING: not every choice works. Correct way:

$$g_i = \sqrt{\mathrm{diag}(p^{(i)})}\, L,$$

L lower triangular Cholesky factor $L^\dagger L = \rho_Y^{-1}$.

Remark:

It is maintained that $g \cdot Y$ is a unit vector the entire time.

Fixing a marginal

Easy to fix i^{th} marginal: choose g_i such that $g_i \rho_Y^{(i)} g_i^\dagger = \mathrm{diag}(p^{(i)})$.
WARNING: not every choice works. Correct way:

$$g_i = \sqrt{\mathrm{diag}(p^{(i)})}\, L,$$

L lower triangular Cholesky factor $L^\dagger L = \rho_Y^{-1}$.

Remark:

It is maintained that $g \cdot Y$ is a unit vector the entire time.

Fixing a marginal

Easy to fix i^{th} marginal: choose g_i such that $g_i \rho_Y^{(i)} g_i^\dagger = \mathrm{diag}(p^{(i)})$.
WARNING: not every choice works. Correct way:

$$g_i = \sqrt{\mathrm{diag}(p^{(i)})}\, L,$$

L lower triangular Cholesky factor $L^\dagger L = \rho_Y^{-1}$.

Remark:

It is maintained that $g \cdot Y$ is a unit vector the entire time.

Analysis

Proof outline

- **The randomization step:** Success = nonvanishing of a potential function on $g_0 \cdot X$

  If potential function nonvanishing, is in fact bounded below by

$$\text{poly}(\langle X \rangle + \langle p \rangle)$$

- **The triangular scaling steps:** the potential function decreases by $\Omega(\varepsilon^2)$ each step provided marginals are ε-far from targets

Proof outline

- **The randomization step:** Success = nonvanishing of a potential function on $g_0 \cdot X$
  If potential function nonvanishing, is in fact bounded below by

$$\text{poly}(\langle X \rangle + \langle p \rangle)$$

- **The triangular scaling steps:** the potential function decreases by $\Omega(\varepsilon^2)$ each step provided marginals are ε-far from targets

Proof outline

- **The randomization step:** Success = nonvanishing of a potential function on $g_0 \cdot X$
  If potential function nonvanishing, is in fact bounded below by
  $$\mathrm{poly}(\langle X \rangle + \langle p \rangle)$$
- **The triangular scaling steps:** the potential function decreases by $\Omega(\varepsilon^2)$ each step provided marginals are ε-far from targets

Description of the potential functions

First define a modified determinant.

Definition

If b is a lower triangular matrix and α a sequence of real numbers, define

$$\chi_\alpha(b) = \prod_{i=1}^{m} b_{ii}^{\alpha_i}.$$

Throughout the iterations, keep track of the following function:

$$f_{\mathbf{p},Y}(g) := \log \frac{\|g \cdot Y\|^2}{|\chi_{\mathbf{p}}(g)|^2}$$

where $\chi_{\mathbf{p}}(g) = \prod_{i=1}^{d} \chi_{\mathbf{p}^{(i)}}(g_i)$.

Description of the potential functions

First define a modified determinant.

Definition

If b is a lower triangular matrix and α a sequence of real numbers, define

$$\chi_\alpha(b) = \prod_{i=1}^{m} b_{ii}^{\alpha_i}.$$

Throughout the iterations, keep track of the following function:

$$f_{\mathbf{p},Y}(g) := \log \frac{\|g \cdot Y\|^2}{|\chi_{\mathbf{p}}(g)|^2}$$

where $\chi_{\mathbf{p}}(g) = \prod_{i=1}^{d} \chi_{\mathbf{p}^{(i)}}(g_i)$.

Description of the potential functions

First define a modified determinant.

Definition

If b is a lower triangular matrix and α a sequence of real numbers, define

$$\chi_\alpha(b) = \prod_{i=1}^{m} b_{ii}^{\alpha_i}.$$

Throughout the iterations, keep track of the following function:

$$f_{\mathbf{p},Y}(g) := \log \frac{\|g \cdot Y\|^2}{|\chi_{\mathbf{p}}(g)|^2}$$

where $\chi_{\mathbf{p}}(g) = \prod_{i=1}^{d} \chi_{\mathbf{p}^{(i)}}(g_i)$.

Triangular scaling steps

Lemma (Change in potential function)

Let $g(t)$ be the scaling in the t^{th} step. If for some i,

$$\left\| \rho^{(i)}_{g(t)\cdot Y} - \operatorname{diag}(\boldsymbol{p}^{(i)}_{\uparrow}) \right\|_{Tr} > \varepsilon,$$

then $f_{\boldsymbol{p},Y}(g(t+1)) \le f_{\boldsymbol{p},Y}(g(t)) - \Omega(\varepsilon^2)$.

Let

$$B := \{g : g_i \text{ lower triangular }\}.$$

Corollary: if $\inf_{g \in B} f_{\boldsymbol{p},Y}(b) = -C$, then the number of iterations is at most

$$O(C/\varepsilon^2).$$

Triangular scaling steps

Lemma (Change in potential function)

Let $g(t)$ be the scaling in the t^{th} step. If for some i,

$$\left\| \rho^{(i)}_{g(t)\cdot Y} - \mathrm{diag}(p^{(i)}_{\uparrow}) \right\|_{Tr} > \varepsilon,$$

then $f_{\mathbf{p},Y}(g(t+1)) \leq f_{\mathbf{p},Y}(g(t)) - \Omega(\varepsilon^2)$.

Let

$$B := \{g : g_i \text{ lower triangular } \}.$$

Corollary: if $\inf_{g \in B} f_{\mathbf{p},Y}(b) = -C$, then the number of iterations is at most

$$O(C/\varepsilon^2).$$

Triangular scaling steps

Lemma (Change in potential function)

Let $g(t)$ be the scaling in the t^{th} step. If for some i,

$$\left\| \rho^{(i)}_{g(t)\cdot Y} - \mathrm{diag}(p^{(i)}_{\uparrow}) \right\|_{Tr} > \varepsilon,$$

then $f_{\mathbf{p},Y}(g(t+1)) \leq f_{\mathbf{p},Y}(g(t)) - \Omega(\varepsilon^2)$.

Let

$$B := \{g : g_i \text{ lower triangular }\}.$$

Corollary: if $\inf_{g\in B} f_{\mathbf{p},Y}(b) = -C$, then the number of iterations is at most

$$O(C/\varepsilon^2).$$

Why does f decrease at all?

Consider

$$f_{\boldsymbol{p},Y}(g(t+1)) - f_{\boldsymbol{p},Y}(g(t))$$

At each step, $\|g(t) \cdot X\|$ is 1, so this is

$$-\log|\chi_{\boldsymbol{p}}(g(t+1))|^2 + \log|\chi_{\boldsymbol{p}}(g(t))|^2$$

Recall that only the i^{th} factor changed was multiplied by $\sqrt{\operatorname{diag}(\boldsymbol{p}^{(i)})}L$, so the above is

$$-\sum_{j=1}^{n_i} \boldsymbol{p}_j^{(i)} \log\left(\boldsymbol{p}_j^{(i)}|L_{jj}|^2\right).$$

However, $\sum_j |L_{jj}|^{-2} \leq \|L^{-1}\|_F = \operatorname{Tr}\rho^{(i)}_{g(t)\cdot Y} = 1$, so the above is

$$-D_{KL}(\boldsymbol{p}^{(i)}||\boldsymbol{q}) < 0.$$

for some (subnormalized) distribution $\boldsymbol{q}$.

Why does f decrease at all?

Consider

$$f_{\mathbf{p},Y}(g(t+1)) - f_{\mathbf{p},Y}(g(t))$$

At each step, $\|g(t)\cdot X\|$ is 1, so this is

$$-\log|\chi_{\mathbf{p}}(g(t+1))|^2 + \log|\chi_{\mathbf{p}}(g(t))|^2$$

Recall that only the i^{th} factor changed was multiplied by $\sqrt{\operatorname{diag}(\mathbf{p}^{(i)})}L$, so the above is

$$-\sum_{j=1}^{n_i} p_j^{(i)} \log\left(p_j^{(i)}|L_{jj}|^2\right).$$

However, $\sum_j |L_{jj}|^{-2} \leq \|L^{-1}\|_F = \operatorname{Tr}\rho^{(i)}_{g(t)\cdot Y} = 1$, so the above is

$$-D_{KL}(\mathbf{p}^{(i)}||\mathbf{q}) < 0.$$

for some (subnormalized) distribution $\mathbf{q}$.

Why does f decrease at all?

Consider

$$f_{\mathbf{p},Y}(g(t+1)) - f_{\mathbf{p},Y}(g(t))$$

At each step, $\|g(t)\cdot X\|$ is 1, so this is

$$-\log|\chi_{\mathbf{p}}(g(t+1))|^2 + \log|\chi_{\mathbf{p}}(g(t))|^2$$

Recall that only the i^{th} factor changed was multiplied by $\sqrt{\mathrm{diag}(\mathbf{p}^{(i)})}L$, so the above is

$$-\sum_{j=1}^{n_i} p_j^{(i)} \log\left(p_j^{(i)}|L_{jj}|^2\right).$$

However, $\sum_j |L_{jj}|^{-2} \le \|L^{-1}\|_F = \operatorname{Tr}\rho^{(i)}_{g(t)\cdot Y} = 1$, so the above is

$$-D_{KL}(\mathbf{p}^{(i)}||\mathbf{q}) < 0.$$

for some (subnormalized) distribution $\mathbf{q}$.

Why does f decrease at all?

Consider

$$f_{\mathbf{p},Y}(g(t+1)) - f_{\mathbf{p},Y}(g(t))$$

At each step, $\|g(t) \cdot X\|$ is 1, so this is

$$-\log|\chi_{\mathbf{p}}(g(t+1))|^2 + \log|\chi_{\mathbf{p}}(g(t))|^2$$

Recall that only the i^{th} factor changed was multiplied by $\sqrt{\operatorname{diag}(\mathbf{p}^{(i)})}L$, so the above is

$$-\sum_{j=1}^{n_i} \mathbf{p}_j^{(i)} \log\left(\mathbf{p}_j^{(i)}|L_{jj}|^2\right).$$

However, $\sum_j |L_{jj}|^{-2} \leq \|L^{-1}\|_F = \operatorname{Tr}\rho^{(i)}_{g(t)\cdot Y} = 1$, so the above is

$$-D_{KL}(\mathbf{p}^{(i)}||\mathbf{q}) < 0.$$

for some (subnormalized) distribution $\mathbf{q}$.

Why does f decrease at all?

Consider

$$f_{\mathbf{p},Y}(g(t+1)) - f_{\mathbf{p},Y}(g(t))$$

At each step, $\|g(t)\cdot X\|$ is 1, so this is

$$-\log|\chi_{\mathbf{p}}(g(t+1))|^2 + \log|\chi_{\mathbf{p}}(g(t))|^2$$

Recall that only the i^{th} factor changed was multiplied by $\sqrt{\operatorname{diag}(\mathbf{p}^{(i)})}L$, so the above is

$$-\sum_{j=1}^{n_i} \mathbf{p}_j^{(i)} \log\left(\mathbf{p}_j^{(i)}|L_{jj}|^2\right).$$

However, $\sum_j |L_{jj}|^{-2} \leq \|L^{-1}\|_F = \operatorname{Tr} \rho^{(i)}_{g(t)\cdot Y} = 1$, so the above is

$$-D_{KL}(\mathbf{p}^{(i)}||\mathbf{q}) < 0.$$

for some (subnormalized) distribution $\mathbf{q}$.

Randomization step: highest weights

A polynomial P on $\mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$ is a highest weight** of weight $\boldsymbol{\lambda}$ if

$$P(g \cdot X) = \chi_{\boldsymbol{\lambda}}(g) P(X)$$

for all $g \in B$. That is, p is a common eigenvector of the action of the lower triangular matrices on the polynomials.

**Really this is a lowest weight of $-\boldsymbol{\lambda}$

Randomization step: highest weights

A polynomial P on $\mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$ is a highest weight** of weight $\boldsymbol{\lambda}$ if

$$P(g \cdot X) = \chi_{\boldsymbol{\lambda}}(g) P(X)$$

for all $g \in B$. That is, p is a common eigenvector of the action of the lower triangular matrices on the polynomials.

**Really this is a lowest weight of $-\boldsymbol{\lambda}$

Randomization step: highest weights

A polynomial P on $\mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$ is a highest weight** of weight $\boldsymbol{\lambda}$ if

$$P(g \cdot X) = \chi_{\boldsymbol{\lambda}}(g) P(X)$$

for all $g \in B$. That is, p is a common eigenvector of the action of the lower triangular matrices on the polynomials.
**Really this is a lowest weight of $-\boldsymbol{\lambda}$

Nonvanishing highest weights = lower bounds

Suppose P is a highest weight of degree k of weight $\boldsymbol{\lambda}$ satisfying $P(Y) :\leq \|P\|\|Y\|^k$ for all Y. Then

$$f_{\boldsymbol{p},Y}(g) \geq \frac{1}{k} \log \frac{|P(Y)|^2}{\|P\|^2}.$$

Proof.

$$kf_{\boldsymbol{p},Y}(g) = \log \frac{\|g \cdot Y\|^{2k}}{|\chi_{\boldsymbol{\lambda}}(g)|^2} \geq \log \frac{|P(g \cdot Y)|^2}{\|P\|^2 |\chi_{\boldsymbol{\lambda}}(g)|^2} = \log \frac{|P(X)|^2}{\|P\|^2}$$

□

Thus, highest weights that do not vanish on Y give us lower bounds!

Nonvanishing highest weights = lower bounds

Suppose P is a highest weight of degree k of weight $\boldsymbol{\lambda}$ satisfying $P(Y) :\leq \|P\| \|Y\|^k$ for all Y. Then

$$f_{\boldsymbol{p},Y}(g) \geq \frac{1}{k} \log \frac{|P(Y)|^2}{\|P\|^2}.$$

Proof.

$$k f_{\boldsymbol{p},Y}(g) = \log \frac{\|g \cdot Y\|^{2k}}{|\chi_{\boldsymbol{\lambda}}(g)|^2} \geq \log \frac{|P(g \cdot Y)|^2}{\|P\|^2 |\chi_{\boldsymbol{\lambda}}(g)|^2} = \log \frac{|P(X)|^2}{\|P\|^2}$$

□

Thus, highest weights that do not vanish on Y give us lower bounds!

Nonvanishing highest weights = lower bounds

Suppose P is a highest weight of degree k of weight $\boldsymbol{\lambda}$ satisfying $P(Y) :\leq \|P\|\|Y\|^k$ for all Y. Then

$$f_{\mathbf{p},Y}(g) \geq \frac{1}{k} \log \frac{|P(Y)|^2}{\|P\|^2}.$$

Proof.

$$kf_{\mathbf{p},Y}(g) = \log \frac{\|g \cdot Y\|^{2k}}{|\chi_{\boldsymbol{\lambda}}(g)|^2} \geq \log \frac{|P(g \cdot Y)|^2}{\|P\|^2 |\chi_{\boldsymbol{\lambda}}(g)|^2} = \log \frac{|P(X)|^2}{\|P\|^2}$$

□

Thus, highest weights that do not vanish on Y give us lower bounds!

Nonzero highest weight after randomization

Theorem (Ness-Mumford '84, Brion '87)

$\boldsymbol{p} \in \Delta(X) \cap \mathbb{Q}$ if and only if some there is some integer ℓ such that $\boldsymbol{\lambda} = \ell\boldsymbol{p}$ is integral and some highest weight $P_{\boldsymbol{\lambda}}$ does not vanish on $\overline{G \cdot X}$.

Further, (Derksen '01), if $\ell\boldsymbol{p}$ is integral we may take $k = (\ell d \max n_i)^{(d \max n_i^2)}$;

Thus in randomization step we may take $K = 2(\ell d \max n_i)^{(d \max n_i^2)}$ to obtain a nonvanishing highest weight with probability 1/2.

Nonzero highest weight after randomization

Theorem (Ness-Mumford '84, Brion '87)

$\boldsymbol{p} \in \Delta(X) \cap \mathbb{Q}$ if and only if some there is some integer ℓ such that $\boldsymbol{\lambda} = \ell \boldsymbol{p}$ is integral and some highest weight $P_{\boldsymbol{\lambda}}$ does not vanish on $\overline{G \cdot X}$.

Further, (Derksen '01), if $\ell \boldsymbol{p}$ is integral we may take $k = (\ell d \max n_i)^{(d \max n_i^2)}$;
Thus in randomization step we may take $K = 2(\ell d \max n_i)^{(d \max n_i^2)}$ to obtain a nonvanishing highest weight with probability 1/2.

Nonvanishing highest weights $\Longrightarrow$ bounded highest weights

Lemma (BFGOWW'18)

The space of highest weights of weight $\boldsymbol{\lambda}$ are spanned by polynomials with integer coefficients and $\|P\| \leq n^k$.

Suppose the largest entry of $g_0 \cdot X$ is M.

Corollary

If a highest weight of $\boldsymbol{\lambda}$ doesn't vanish on $g_0 \cdot X$, then $\inf_{g \in B} f_{\mathbf{p},Y}(g) \geq -2\log n - \log \|g_0 \cdot X\|^2 \geq -3\log n - \log M$

Corollary

The algorithm runs in $O((\log n + \log M)/\varepsilon^2) = O(d \max n_i(\langle X \rangle + \langle \mathbf{p} \rangle + \log d \max n_i)/\varepsilon^2)$ *steps*

Nonvanishing highest weights $\implies$ bounded highest weights

Lemma (BFGOWW'18)

The space of highest weights of weight $\boldsymbol{\lambda}$ are spanned by polynomials with integer coefficients and $\|P\| \leq n^k$.

Suppose the largest entry of $g_0 \cdot X$ is M.

Corollary

If a highest weight of $\boldsymbol{\lambda}$ doesn't vanish on $g_0 \cdot X$, then $\inf_{g \in B} f_{\mathbf{p},Y}(g) \geq -2\log n - \log\|g_0 \cdot X\|^2 \geq -3\log n - \log M$

Corollary

The algorithm runs in $O((\log n + \log M)/\varepsilon^2) = O(d \max n_i(\langle X \rangle + \langle \mathbf{p} \rangle + \log d \max n_i)/\varepsilon^2)$ *steps*

Nonvanishing highest weights $\implies$ bounded highest weights

Lemma (BFGOWW'18)

The space of highest weights of weight $\boldsymbol{\lambda}$ are spanned by polynomials with integer coefficients and $\|P\| \leq n^k$.

Suppose the largest entry of $g_0 \cdot X$ is M.

Corollary

If a highest weight of $\boldsymbol{\lambda}$ doesn't vanish on $g_0 \cdot X$, then $\inf_{g \in B} f_{\boldsymbol{p},Y}(g) \geq -2\log n - \log \|g_0 \cdot X\|^2 \geq -3\log n - \log M$

Corollary

The algorithm runs in $O((\log n + \log M)/\varepsilon^2) = O(d \max n_i(\langle X \rangle + \langle \boldsymbol{p} \rangle + \log d \max n_i)/\varepsilon^2)$ *steps*

Nonvanishing highest weights $\Longrightarrow$ bounded highest weights

Lemma (BFGOWW'18)

The space of highest weights of weight $\boldsymbol{\lambda}$ are spanned by polynomials with integer coefficients and $\|P\| \leq n^k$.

Suppose the largest entry of $g_0 \cdot X$ is M.

Corollary

If a highest weight of $\boldsymbol{\lambda}$ doesn't vanish on $g_0 \cdot X$, then $\inf_{g \in B} f_{\mathbf{p},Y}(g) \geq -2 \log n - \log \|g_0 \cdot X\|^2 \geq -3 \log n - \log M$

Corollary

The algorithm runs in $O((\log n + \log M)/\varepsilon^2) = O(d \max n_i(\langle X \rangle + \langle \mathbf{p} \rangle + \log d \max n_i)/\varepsilon^2)$ *steps*

Open problems

Open problems

- Obtain $\operatorname{poly}\log(1/\varepsilon)$ run time!
- Solve the optimization problem for other group actions (in progress).
- Develop separation oracles for moment polytopes.

Thank you!

Moment polytopes

Suppose G acts linearly on a vector space V and the inner product $\langle -, - \rangle$ is invariant under the unitaries $K = U(n_1) \times \cdots \times U(n_d)$.

Definition

The map $\mu : V \to \mathrm{Herm}_{n_1} \times \cdots \times \mathrm{Herm}_{n_d}$ given by

$$\mu : X \mapsto \nabla_{A=0} \log \|e^A \cdot X\|$$

is known as the moment map for the action of G on V.

General framework

Suppose G acts linearly on a vector space V and the inner product $\langle -, - \rangle$ is invariant under the unitaries $K = U(n_1) \times \cdots \times U(n_d)$.

Definition

The map $\mu : V \to \mathbf{Herm}_{n_1} \times \cdots \times \mathbf{Herm}_{n_d}$ given by

$$\mu : X \mapsto \nabla_{A=0} \log \|e^A \cdot X\|$$

is known as the moment map for the action of G on V.

Moment polytope

Define

$$\Delta(X) = \left\{(\mathrm{spec}(\mu^{(1)}(Y)), \dots, \mathrm{spec}(\mu^{(d)}(Y)) : Y \in \overline{G \cdot X}\right\}.$$

Amazingly, $\Delta(X)$ is not only a polytope but encodes the rep. theory of polynomials on $\overline{G \cdot X}$!

G acts on a polynomial p on V by $g \cdot p(x) = p(g^{-1} \cdot x)$.

Theorem (Mumford '84, Brion '87)

$$\Delta(X) \cap \mathbb{Q} = \{\boldsymbol{\lambda}/k : V_{G,\lambda} \subset R_k(\overline{G \cdot X})\}$$

Moment polytope

Define

$$\Delta(X) = \left\{ (\mathrm{spec}(\mu^{(1)}(Y)), \ldots, \mathrm{spec}(\mu^{(d)}(Y)) : Y \in \overline{G \cdot X} \right\}.$$

Amazingly, $\Delta(X)$ is not only a polytope but encodes the rep. theory of polynomials on $\overline{G \cdot X}$!

G acts on a polynomial p on V by $g \cdot p(x) = p(g^{-1} \cdot x)$.

Theorem (Mumford '84, Brion '87)

$$\Delta(X) \cap \mathbb{Q} = \{\boldsymbol{\lambda}/k : V_{G,\lambda} \subset R_k(\overline{G \cdot X})\}$$

Moment polytope

Define

$$\Delta(X) = \left\{ (\mathrm{spec}(\mu^{(1)}(Y)), \ldots, \mathrm{spec}(\mu^{(d)}(Y)) : Y \in \overline{G \cdot X} \right\}.$$

Amazingly, $\Delta(X)$ is not only a polytope but encodes the rep. theory of polynomials on $\overline{G \cdot X}$!
G acts on a polynomial p on V by $g \cdot p(x) = p(g^{-1} \cdot x)$.

Theorem (Mumford '84, Brion '87)

$$\Delta(X) \cap \mathbb{Q} = \{ \boldsymbol{\lambda}/k : V_{G,\lambda} \subset R_k(\overline{G \cdot X}) \}$$

Moment polytope

Define

$$\Delta(X) = \left\{ (\mathrm{spec}(\mu^{(1)}(Y)), \ldots, \mathrm{spec}(\mu^{(d)}(Y)) : Y \in \overline{G \cdot X} \right\}.$$

Amazingly, $\Delta(X)$ is not only a polytope but encodes the rep. theory of polynomials on $\overline{G \cdot X}$!
G acts on a polynomial p on V by $g \cdot p(x) = p(g^{-1} \cdot x)$.

Theorem (Mumford '84, Brion '87)

$$\Delta(X) \cap \mathbb{Q} = \{\boldsymbol{\lambda}/k : V_{G,\lambda} \subset R_k(\overline{G \cdot X})\}$$