Quantum marginals, invariants, and non-commutative optimization

Michael Walter

University of Amsterdam · QuSoft (Research Center for Quantum Software) · NWO

Leiden, May 2019

based on joint work with Bürgisser, Franks, Garg, Oliveira, Wigderson
(ITCS'18, FOCS'18, arXiv:1905.xxxxx)

Outline and philosophy

Marginal & scaling problems $\longleftrightarrow$ Null cone problems

(Geometry) (Invariant theory)

Interesting class of problems — with applications in q. information, computer science, algebra, analysis — that *surprisingly* can be phrased as optimization problems over noncommutative groups.

Result: General framework and algorithms for this class.

Plan: Introduction & illustration via quantum marginal problem.

Philosophy: An old duality in geometric invariant theory leads to new optimization algorithms.

Example: Matrix scaling

Let X be matrix with nonnegative entries. A *scaling* of X is a matrix

$$Y = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} X \begin{pmatrix} b_1 & & \\ & \ddots & \\ & & b_n \end{pmatrix} \qquad (a_1, \ldots, b_n > 0).$$

A matrix is called *doubly stochastic (d.s.)* if row & column sums are 1.

Matrix scaling (Geometry): Given X, $\exists$ (approximately) d.s. scalings?

Permanent (Invariant Theory): ...iff $\operatorname{per}(X) > 0$!

- can be decided in polynomial time
- find scalings by alternatingly fixing rows & columns ☺ [Sinkhorn]

Connections to complexity, combinatorics, geometry, numerics, ...

Example: Matrix scaling

Let X be matrix with nonnegative entries. A *scaling* of X is a matrix

$$Y = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} X \begin{pmatrix} b_1 & & \\ & \ddots & \\ & & b_n \end{pmatrix} \qquad (a_1, \dots, b_n > 0).$$

A matrix is called *doubly stochastic (d.s.)* if row & column sums are 1.

Matrix scaling (Geometry): Given X, $\exists$ (approximately) d.s. scalings?

Permanent (Invariant Theory): ...iff $\mathrm{per}(X) > 0$!

- can be decided in polynomial time
- find scalings by alternatingly fixing rows & columns ☺ [Sinkhorn]

Connections to complexity, combinatorics, geometry, numerics, ...

Example: Sinkhorn algorithm

$$\begin{pmatrix} 1 & 2 \\ 4 & 0 \end{pmatrix} \xrightarrow{\text{fix rows}} \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ 1 & 0 \end{pmatrix} \xrightarrow{\text{fix cols}} \begin{pmatrix} \frac{1}{4} & 1 \\ \frac{3}{4} & 0 \end{pmatrix} \longrightarrow \dots \longrightarrow \begin{pmatrix} \frac{1}{2t} & 1 \\ \frac{2t-1}{2t} & 0 \end{pmatrix}$$

after t steps. Why does it work? Permanent of $X/\sum_{i,j} X_{ij}$ increases monotonically – can be used to control convergence:

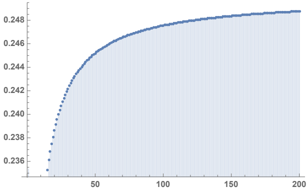
permanent distance to doubly stochastic

State-of-the-art algorithms directly optimize $(a, b) \mapsto \operatorname{per}(aXb)$.

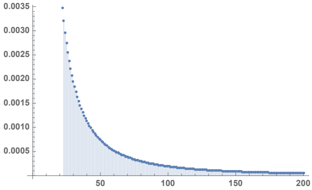
Example: Sinkhorn algorithm

$$\begin{pmatrix}1 & 2\\ 4 & 0\end{pmatrix} \xrightarrow{\text{fix rows}} \begin{pmatrix}\frac{1}{3} & \frac{2}{3}\\ 1 & 0\end{pmatrix} \xrightarrow{\text{fix cols}} \begin{pmatrix}\frac{1}{4} & 1\\ \frac{3}{4} & 0\end{pmatrix} \longrightarrow \dots \longrightarrow \begin{pmatrix}\frac{1}{2t} & 1\\ \frac{2t-1}{2t} & 0\end{pmatrix}$$

after t steps. Why does it work? Permanent of $X/\sum_{i,j} X_{ij}$ increases monotonically – can be used to control convergence:



permanent



distance to doubly stochastic

State-of-the-art algorithms directly optimize $(a, b) \mapsto \operatorname{per}(aXb)$.

Example: Operator scaling and non-commutative PIT

Let $T(\rho) = \sum_i X_i \rho X_i^\dagger$ be a CP map. A *scaling* of T is of the form

$$S = AT(B \cdot B^\dagger)A^\dagger.$$

A map is *unital (U)* if $T(I) = I$ and *trace-preserving (TP)* if $T^\dagger(I) = I$.

Operator scaling (Geometry): Given T, ∃ (approx.) U & TP scalings?

Non-commutative PIT (Invariant Theory): ...iff symbolic matrix $\sum_i y_i X_i$ in *non-commutative* variables y_i is invertible.

- can be decided in polynomial time [Garg et al, Ivanyos et al]
- find scalings by alternatingly making the map U or TP ☺ [Gurvits]

Many further characterizations ($\exists Y_i : \det \sum_i Y_i \otimes X_i \neq 0$) & connections (Brascamp-Lieb inequalities, Paulsen problem, ...).

Example: Operator scaling and non-commutative PIT

Let $T(\rho) = \sum_i X_i \rho X_i^\dagger$ be a CP map. A *scaling* of T is of the form

$$S = AT(B \cdot B^\dagger)A^\dagger.$$

A map is *unital (U)* if $T(I) = I$ and *trace-preserving (TP)* if $T^\dagger(I) = I$.

Operator scaling (Geometry): Given T, $\exists$ (approx.) U & TP scalings?

Non-commutative PIT (Invariant Theory): ...iff symbolic matrix $\sum_i y_i X_i$ in *non-commutative* variables y_i is invertible.

- can be decided in polynomial time [Garg et al, Ivanyos et al]
- find scalings by alternatingly making the map U or TP ☺ [Gurvits]

Many further characterizations ($\exists Y_i : \det \sum_i Y_i \otimes X_i \neq 0$) & connections (Brascamp-Lieb inequalities, Paulsen problem, ...).

Example: Horn problem

Let $\alpha_1 \geq \ldots \geq \alpha_n \geq 0$, $\beta_1 \geq \ldots \geq \beta_n \geq 0$, $\gamma_1 \geq \ldots \geq \gamma_n \geq 0$ be integers.

Horn problem (Geometry): When $\exists$ Hermitian $n \times n$ matrices A, B, C with spectrum α, β, γ such that $A + B = C$?

- Horn conjectured linear inequalities on α, β, γ.

Saturation property (Invariant theory): ...iff *Littlewood-Richardson coefficient* $c^{\gamma}_{\alpha,\beta} > 0$ [Knutson-Tao]

- Horn inequalities sufficient
- lead to *only known* poly-time algorithm [Mulmuley]
- can find A, B, C by natural iterative algorithm [Franks]

All these examples are special cases of a general class of problems. We now focus on 'representative' example involving quantum states!

Example: Horn problem

Let $\alpha_1 \geq \ldots \geq \alpha_n \geq 0$, $\beta_1 \geq \ldots \geq \beta_n \geq 0$, $\gamma_1 \geq \ldots \geq \gamma_n \geq 0$ be integers.

Horn problem (Geometry): When $\exists$ Hermitian $n \times n$ matrices A, B, C with spectrum α, β, γ such that $A + B = C$?

- Horn conjectured linear inequalities on α, β, γ.

Saturation property (Invariant theory): ...iff *Littlewood-Richardson coefficient* $c_{\alpha,\beta}^{\gamma} > 0$ [Knutson-Tao]

- Horn inequalities sufficient
- lead to *only known* poly-time algorithm [Mulmuley]
- can find A, B, C by natural iterative algorithm [Franks]

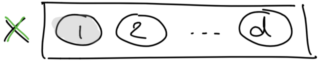
All these examples are special cases of a general class of problems. We now focus on 'representative' example involving quantum states!

Geometry: Quantum states and marginals

Quantum state of d particles is described by unit vector

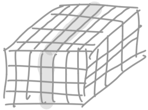
$$X \in V = (\mathbb{C}^n)^{\otimes d} = \mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n$$

$$\rightsquigarrow [X] = |X\rangle \langle X| \in \mathbb{P}(V)$$



State of individual particles described by density matrices $\rho_1^X, \ldots, \rho_d^X$:

$$\mathrm{tr}[\rho_1^X H_1] = \langle (H_1 \otimes I \otimes \ldots \otimes I) X, X \rangle \quad \forall H_1$$



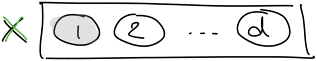
Quantum marginal problem: Which $\rho_1, \ldots, \rho_d$ are consistent with a global state X?

Geometry: Quantum states and marginals

Quantum state of d particles is described by unit vector

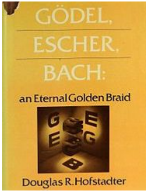
$$X \in V = (\mathbb{C}^n)^{\otimes d} = \mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n$$

$$\rightsquigarrow [X] = |X\rangle\langle X| \in \mathbb{P}(V)$$



State of individual particles described by density matrices $\rho_1^X, \ldots, \rho_d^X$:

$$\operatorname{tr}[\rho_1^X H_1] = \langle (H_1 \otimes I \otimes \ldots \otimes I) X, X \rangle \quad \forall H_1$$



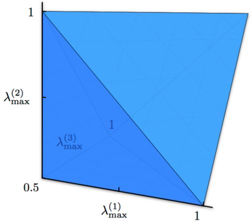
Quantum marginal problem: Which $\rho_1, \ldots, \rho_d$ are consistent with a global state X?

Examples

Two particles: ρ_A and ρ_B compatible with global pure state iff same nonzero eigenvalues (Schmidt decomposition)

Three particles:

$$\lambda_{A,\max} + \lambda_{B,\max} \leq \lambda_{C,\max} + 1$$
$$\lambda_{A,\max} + \lambda_{C,\max} \leq \lambda_{B,\max} + 1$$
$$\lambda_{B,\max} + \lambda_{C,\max} \leq \lambda_{A,\max} + 1$$



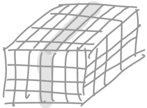
[Higuchi, Sudbery, Szulc]

- necessary and sufficient for qubits
- follows from variational principle: $\lambda_{A,\max} = \max_{\phi_A} \langle \phi_A | \rho_A | \phi_A \rangle$ etc.

Tensor scaling and SLOCC

$X \in V = (\mathbb{C}^n)^{\otimes d}$

$G = \mathrm{SL}(n)^d$ acts on $V = (\mathbb{C}^n)^{\otimes d}$ by $X \mapsto (A_1 \otimes \ldots \otimes A_d) X$



Group orbit = tensor scalings = states that can be obtained by SLOCC (postselected local operations & classical communication).

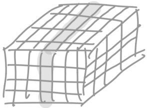
Tensor scaling problem: Which $\rho_1, \ldots, \rho_d$ arise from scaling of given X?

- X fixes the entanglement class
- e.g., for $\rho_i \propto I$, each system maximally entangled with rest
  (quantum version of stochastic tensor)
- in general, answer only depends on eigenvalues λ_i of ρ_i

Tensor scaling and SLOCC

$X \in V = (\mathbb{C}^n)^{\otimes d}$

$G = \mathrm{SL}(n)^d$ acts on $V = (\mathbb{C}^n)^{\otimes d}$ by $X \mapsto (A_1 \otimes \ldots \otimes A_d) X$



Group orbit = tensor scalings = states that can be obtained by SLOCC (postselected local operations & classical communication).

Tensor scaling problem: Which $\rho_1, \ldots, \rho_d$ arise from scaling of given X?

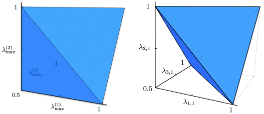
- X fixes the entanglement class
- e.g., for $\rho_i \propto I$, each system maximally entangled with rest
  (quantum version of stochastic tensor)
- in general, answer only depends on eigenvalues λ_i of ρ_i

Tensor scaling and entanglement polytopes

Thus, answer to tensor scaling problem for X is encoded by:

$$\Delta(X) = \Big\{(\lambda_1,\ldots,\lambda_d) \text{ for scalings of } X \text{ (and limits)}\Big\} \subseteq \mathbb{R}^{dn}$$

e.g., for three qubits, $GHZ = |000\rangle + |111\rangle$ *and* $W = |100\rangle + |010\rangle + |001\rangle$*:*



In general:

- convex polytopes [Kirwan, Mumford, W-Christandl-Doran-Gross, Sawicki-Oszmaniec-Kus]
- encode all local info about entanglement class ('entanglement polytopes')
- descriptions by vertices or inequalities intractable (when known)

[Berenstein-Sjamaar, Klyachko, Ressayre, Vergne-W.]

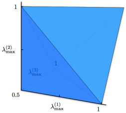
We provide algorithmic solution!

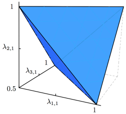
Tensor scaling and entanglement polytopes

Thus, answer to tensor scaling problem for X is encoded by:

$$\Delta(X) = \Big\{(\lambda_1, \ldots, \lambda_d) \text{ for scalings of } X \text{ (and limits)}\Big\} \subseteq \mathbb{R}^{dn}$$

e.g., for three qubits, $GHZ = |000\rangle + |111\rangle$ *and* $W = |100\rangle + |010\rangle + |001\rangle$*:*





In general:

- convex polytopes [Kirwan, Mumford, W-Christandl-Doran-Gross, Sawicki-Oszmaniec-Kus]
- encode all local info about entanglement class ('entanglement polytopes')
- descriptions by vertices or inequalities intractable (when known)

[Berenstein-Sjamaar, Klyachko, Ressayre, Vergne-W.]

We provide algorithmic solution!

The Algorithm

Given λ_A, λ_B, λ_C and reference state X, want $Y = (A \otimes B \otimes C) X$ with these marginals. For simplicity, uniform marginals ($\lambda_A \propto 1_A$ etc).

Algorithm: Start with $Y = X$. For $t = 1, \ldots, T$:
Compute marginals ρ_A, ρ_B, ρ_C of Y. If ε-close to uniform, stop.
Otherwise, replace Y by $e^{-c(\rho_A^o + \rho_B^o + \rho_C^o)} Y$.   *$X^o$ = traceless part*

Result

Algorithm finds $Y = (A \otimes B \otimes C) X$ with marginals ε-close to uniform within $T = \text{poly}(\frac{1}{\varepsilon}, \text{input size})$ steps.

- also works for bosons, fermions, $d > 3$ subsystems, MPS, ...
- can run on quantum computer (but how well? ☺)
- solve quantum marginal problem by using random X

cf. algorithm by Verstraete et al (w/o rigorous analysis)

Why does it work?

"Otherwise, replace X by $e^{-c(\rho_A^o+\rho_B^o+\rho_C^o)}X$."

This step implements gradient descent for the function

$$N(A,B,C) = \|(A \otimes B \otimes C) X\|^2$$

where A,B,C have det=1. Indeed, for traceless $H_A,\dots,H_C$:

$$\frac{1}{2}\partial_{t=0} N(e^{tH_A}, e^{tH_B}, e^{tH_C}) = \mathrm{tr}[\rho_A^o H_A] + \mathrm{tr}[\rho_B^o H_B] + \mathrm{tr}[\rho_C^o H_C],$$

so gradient can be identified with $\rho_A^o, \rho_B^o, \rho_C^o$. Moreover:

- gradient vanishes iff marginals uniform ☺
- log-convexity: $\partial_t^2 \geq 0$, so critical points are global minima ☺

Hold on...

Why does it work?

"Otherwise, replace X by $e^{-c(\rho_A^o+\rho_B^o+\rho_C^o)}X$."

This step implements gradient descent for the function

$$N(A,B,C) = \|(A\otimes B\otimes C)X\|^2$$

where A,B,C have det=1. Indeed, for traceless $H_A,\ldots,H_C$:

$$\frac{1}{2}\partial_{t=0}N(e^{tH_A},e^{tH_B},e^{tH_C}) = \mathrm{tr}[\rho_A^o H_A] + \mathrm{tr}[\rho_B^o H_B] + \mathrm{tr}[\rho_C^o H_C],$$

so gradient can be identified with $\rho_A^o, \rho_B^o, \rho_C^o$. Moreover:

- gradient vanishes iff marginals uniform ☺
- log-convexity: $\partial_t^2 \geq 0$, so critical points are global minima ☺

Hold on...

Why does it work?

"Otherwise, replace X by $e^{-c(\rho_A^o+\rho_B^o+\rho_C^o)}X$."

This step implements gradient descent for the function

$$N(A,B,C) = \|(A\otimes B\otimes C)X\|^2$$

where A,B,C have det=1. Indeed, for traceless H_A,...,H_C:

$$\frac{1}{2}\partial_{t=0}N(e^{tH_A},e^{tH_B},e^{tH_C}) = \mathrm{tr}[\rho_A^o H_A] + \mathrm{tr}[\rho_B^o H_B] + \mathrm{tr}[\rho_C^o H_C],$$

so gradient can be identified with $\rho_A^o, \rho_B^o, \rho_C^o$. Moreover:

- gradient vanishes iff marginals uniform ☺
- log-convexity: $\partial_t^2 \geq 0$, so critical points are global minima ☺

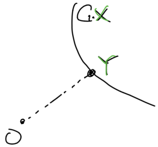
Hold on...

Non-commutative duality $G = \mathrm{SL}(n)^d$

The following optimization problems are equivalent: [Kempf-Ness]

$$\boxed{\inf_{g \in G} \|g \cdot X\| > 0} \iff \boxed{\inf_{g \in G} \mathrm{ds}(g \cdot X) = 0}$$

$$\mathrm{ds}(Y) := \sum_{i=1}^{d} \|\rho_i^Y - \tfrac{I}{n}\|^2$$



- primal: norm minimization, dual: marginal problem
- non-commutative version of LP duality

We develop general duality theory and 1st & 2nd order methods.

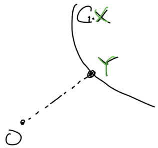
All examples from introduction fall into this framework!

Everything works for general actions of reductive G. Primal is log-convex along geodesics.

Invariant theory

$G = \mathrm{SL}(n)^d$ acts on $V = (\mathbb{C}^n)^{\otimes d}$, so also on ring of polynomials.

Primal problem (norm minimization) is equivalent to classical problem in invariant theory:



Null cone problem: Given X, $\exists$ G-invariant poly P s.th. $P(X) \neq P(0)$?

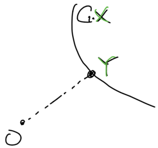
- even interesting for generic X: existence of invariants
  (in general, NP-hard for fixed degree)
- using standard algorithms infeasible already for small d, n

Numerical algorithm solves an algebraic problem! Conversely, we use invariant theory in analysis of algorithm...

Invariant theory

$G = \mathrm{SL}(n)^d$ acts on $V = (\mathbb{C}^n)^{\otimes d}$, so also on ring of polynomials.

Primal problem (norm minimization) is equivalent to classical problem in invariant theory:



Null cone problem: Given X, $\exists$ G-invariant poly P s.th. $P(X) \neq P(0)$?

- even interesting for generic X: existence of invariants
  (in general, NP-hard for fixed degree)
- using standard algorithms infeasible already for small d, n

Numerical algorithm solves an algebraic problem! Conversely, we use invariant theory in analysis of algorithm...

Analysis of Algorithm

"Unless ε-close to uniform, replace Y by $e^{-c(\rho_A^o+\rho_B^o+\rho_C^o)}Y$."

To obtain rigorous algorithm, show:

- *progress in each step:* $\|e^{-c(\rho_A^o+\rho_B^o+\rho_C^o)}Y\| \leq (1-c_1\varepsilon)\|Y\|$
- *a priori lower bound:* $\inf_{\det=1}\|(A\otimes B\otimes C)X\| \geq c_2$

Then, $(1-c_1\varepsilon)^T \geq c_2$ bounds the number of steps T.

The first point follows from convexity estimates.

For the second, construct 'explicit' invariants with 'nice' coefficients and $P(X)\neq 0$ to obtain bound in terms of bitsize of X.

Analysis of Algorithm

"Unless ε-close to uniform, replace Y by $e^{-c(\rho_A^o+\rho_B^o+\rho_C^o)}Y$."

To obtain rigorous algorithm, show:

- *progress in each step:* $\|e^{-c(\rho_A^o+\rho_B^o+\rho_C^o)}Y\| \leq (1-c_1\varepsilon)\|Y\|$
- *a priori lower bound:* $\inf_{\det=1}\|(A\otimes B\otimes C)X\| \geq c_2$

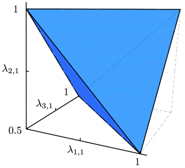
Then, $(1-c_1\varepsilon)^T \geq c_2$ bounds the number of steps T.

The first point follows from convexity estimates.

For the second, construct 'explicit' invariants with 'nice' coefficients and $P(X)\neq 0$ to obtain bound in terms of bitsize of X.

Summary and outlook

arXiv:1804.04739, 1905.xxxxx



Marginal & scaling problems

↕ duality

Norm minimization

Effective algorithms for large class of problems, incl. quantum marginal problem (also fermions) and tensor scaling. Based on geodesically convex optimization and invariant theory.

Many exciting directions:

- Numerical studies in q. many-body systems or chemistry
- Quantum algorithms?
- Algorithms for other problems with natural symmetries?
- *What are the 'tractable' problems in invariant theory?* $\mathbb{C} \rightsquigarrow \mathbb{F}$?

Thank you for your attention!

The tensor scaling algorithm

Input: $X \in V$ rational, $\varepsilon > 0$

- If any ρ_i^X is singular: Null cone ↯
- Set $Y^{(0)} := X$.
- For $t = 0, 1, \ldots, T$:
  - If $\mathrm{ds}(T^{(t)}) < \varepsilon$: Success ☺
  - Choose i such that $\|\rho_i^{Y^{(t)}} - \frac{I}{n}\| > \frac{\varepsilon}{\sqrt{d}}$ and apply tensor scaling step:

$$Y^{(t+1)} \leftarrow (n\rho_i^{Y^{(t)}})^{-1/2} \cdot Y^{(t)}$$

- Null cone ↯

Other target spectra: Adjust tensor scaling step (in particular, use Cholesky square root) and randomize initial point.

A general equivalence

$\mathcal{V} \subseteq \mathbb{P}(V)$

All points in $\Delta(\mathcal{V})$ can be described via invariant theory:

$$V_\lambda \subseteq \mathbb{C}[\mathcal{V}]_{(k)} \quad \Rightarrow \quad \frac{\lambda}{k} \in \Delta(\mathcal{V})$$

(λ highest weight, k degree)

- Can also study multiplicities $g(\lambda, k) := \# V_\lambda \subseteq \mathbb{C}[\mathcal{V}]_{(k)}$.
- This leads to interesting computational problems:

$g = ?$	$g > 0?$	$\exists s > 0 : g(s\lambda, sk) > 0?$
(#-hard)	(NP-hard)	(our problem!)

Completely unlike Horn's problem: *Knutson–Tao saturation property does not hold, and hence we can hope for efficient algorithms!*