Scaling Algorithms for Polynomial Identity Testing

Zeyuan Allen-Zhu
Ankit Garg
Yuanzhi Li
Rafael Oliveira
Avi Wigderson
Today

Scaling Problems
Null Cone
Alternating Minimization
Orbit Closure Intersection
Geodesic Convexity
$n \times n$ complex matrix $A$ is **doubly balanced (DB)** if $\ell_2$ norm of rows/columns of $A$ are equal.

$B$ is **scaling** of $A$ if $\exists$ complex $x_1, \ldots, x_n, y_1, \ldots, y_n$ s.t. $\prod x_i = \prod y_j = 1$ and $b_{ij} = x_i a_{ij} y_j$.

$A$ has DB scaling if $\exists$ scaling $B$ of $A$ s.t. $B$ is DB.

$$db(A) = \sum_i \left( \frac{r_i}{||A||^2} - \frac{1}{n} \right)^2 + \sum_j \left( \frac{c_j}{||A||^2} - \frac{1}{n} \right)^2$$

$A$ has approx. DB scaling if $\forall \epsilon > 0$ there is scaling $B_\epsilon$ of $A$ s.t. $db(B_\epsilon) < \epsilon$.

1. When does $A$ have approx. DB scaling?
2. Can we find it efficiently?
Matrix Balancing - examples

\[
\begin{align*}
\sqrt{2} & \quad (\sqrt{2})^{-1} \\
1 & \quad 1 \\
1 & \quad 4 \\
\end{align*}
\]

\[
\begin{align*}
\sqrt{2} & \quad (\sqrt{2})^{-1} \\
1 & \quad 1 \\
1 & \quad 4 \\
\end{align*}
\]

\[
\begin{align*}
\sqrt{2} & \quad (\sqrt{2})^{-1} \\
1 & \quad 1 \\
1 & \quad 4 \\
\end{align*}
\]

\[
\begin{align*}
\sqrt{2} & \quad (\sqrt{2})^{-1} \\
1 & \quad 1 \\
1 & \quad 4 \\
\end{align*}
\]

\[
\begin{align*}
\sqrt{2} & \quad (\sqrt{2})^{-1} \\
1 & \quad 1 \\
1 & \quad 4 \\
\end{align*}
\]

\[
\begin{align*}
\sqrt{2} & \quad (\sqrt{2})^{-1} \\
1 & \quad 1 \\
1 & \quad 4 \\
\end{align*}
\]

\[
\begin{align*}
\sqrt{2} & \quad (\sqrt{2})^{-1} \\
1 & \quad 1 \\
1 & \quad 4 \\
\end{align*}
\]

\[
\begin{align*}
\sqrt{2} & \quad (\sqrt{2})^{-1} \\
1 & \quad 1 \\
1 & \quad 4 \\
\end{align*}
\]

\[
\begin{align*}
\sqrt{2} & \quad (\sqrt{2})^{-1} \\
1 & \quad 1 \\
1 & \quad 4 \\
\end{align*}
\]

\[
\begin{align*}
\sqrt{2} & \quad (\sqrt{2})^{-1} \\
1 & \quad 1 \\
1 & \quad 4 \\
\end{align*}
\]
Problem: $A \in M_n(\mathbb{C})$, $\epsilon > 0$, is there $\epsilon$-scaling to DB? If yes, find it.

Algorithm S [Kruithof’37, …, Sinkhorn’64]:

Repeat $k$ times:

1. Normalize rows of $A$ (make norm of rows equal)
2. Normalize columns of $A$ (make norm of cols equal)

If at any point $\text{db}(A) < \epsilon$, output the scaling.

Else, output: no scaling.

Questions:

• Are we making progress at all?
• How do we know when to stop? (Which $k$?)
• Is there an $\epsilon_0$ such that if can scale to $\epsilon_0$ then can scale for any $\epsilon$?
Question: How can we distinguish between these two cases?

Observation: In first example, “huge” block of zeros (Hall blocker). In second, have a matching.

Are these the only cases?
A **quantum operator** is any map $T: M_n(\mathbb{C}) \to M_n(\mathbb{C})$ given by $(A_1, \ldots, A_m)$ s.t.

$$T(X) = \sum_{1 \leq i \leq m} A_i X A_i^\dagger$$

Such maps take psd matrices to psd matrices.

Dual of $T(X)$ is map $T^*: M_n(\mathbb{C}) \to M_n(\mathbb{C})$ given by:

$$T^*(X) = \sum_{1 \leq i \leq m} A_i^\dagger X A_i$$

- Analog of scaling?
- Double balanced?

Can scaling solve PIT?
Operator balancing

A quantum operator $T: M_n(\mathbb{C}) \to M_n(\mathbb{C})$ is **doubly balanced (DB)** if $T(I) = T^*(I) = I$.

Scaling of $T(X)$ consists of $L, R \in SL_n(\mathbb{C})$ s.t.

$$(A_1, ..., A_m) \to (LA_1R, ..., LA_mR)$$

Distance to doubly-balanced:

$$db(T) \overset{\text{def}}{=} \left\| \frac{T(I)}{\|A\|^2} - \frac{1}{n} I \right\|_F^2 + \left\| \frac{T^*(I)}{\|A\|^2} - \frac{1}{n} I \right\|_F^2$$

$T(X)$ has approx. DB scaling if $\forall \epsilon > 0$, $\exists$ scaling $L_\epsilon, R_\epsilon$ s.t. operator $T_\epsilon(X)$ given by $(L_\epsilon A_1 R_\epsilon, ..., L_\epsilon A_m R_\epsilon)$ has $db(T_\epsilon) \leq \epsilon$.

1. When does $(A_1, ..., A_m)$ have approx. DB scaling?
2. Can we find it efficiently?
Problem: operator $\mathbf{T} = (A_1, \ldots, A_m)$, $\epsilon > 0$, can $\mathbf{T}$ be $\epsilon$-scaled to double stochastic? If yes, find scaling.

Algorithm G [Gurvits’ 04]:

Repeat $k$ times:
1. Left normalize $\mathbf{T}(\mathbf{X})$, i.e. make $\mathbf{T}(\mathbf{I}) \propto \mathbf{I}$.
2. Right normalize $\mathbf{T}(\mathbf{X})$, i.e. make $\mathbf{T}^*(\mathbf{I}) \propto \mathbf{I}$.

If at any point $\text{dB}(\mathbf{T}) < \epsilon$ output scaling.
Else output no scaling.

• Which $k$ should we choose?
• Is there an $\epsilon_0$ such that if can scale to $\epsilon_0$ then can scale for any $\epsilon$?
Analysis – General Approach

Three steps:

1. **[Upper bound]** Potential function $\Phi$ is norm of input.
   - $\Phi$ upper bounded by input size

2. **[Progress/step]** If we are $\epsilon$-far from DB then normalization decreases value of $\Phi$ by $\times \exp(O(\epsilon))$

3. **[Lower bound]** If there is scaling, “some property” tells us that $\Phi \geq \exp(-\text{poly}(n))$
   - Bounded away from zero

Approach proves correctness & running time of $\text{poly}(nb/\epsilon)$
Analysis – Revisited (matrix scaling)

Three steps:

1. **[Upper bound]** Potential function $\Phi = ||A||^2$
   - $\Phi$ upper bounded by input size

2. **[Progress/step]** If we are $\epsilon$-far from DB then normalization decreases value of $\Phi$ by $\times \exp(O(\epsilon))$

3. **[Lower bound]** $A$ not in null cone, there is “nice” invariant $p(Z)$ s.t. $p(A) \neq 0$.
   - $p(Z)$ invariant $\Rightarrow p(A) = p(B), \ \forall B \in G \cdot A$
   - $p(Z)$ integer coeffs. $\Rightarrow |p(A)|^2 \geq 1 \Rightarrow ||B|| \geq \exp(-n)$

Proves correctness & running time of $\text{poly}(n \cdot \log(\nu)/\epsilon)$
Matrix Scaling: $ST_n \times ST_n \sim M_n(\mathbb{C})$

- Matching monomials are invariants:

$$B = XAY \Rightarrow \prod b_{i\sigma(i)} = \prod x_i a_{i\sigma(i)} y_{\sigma(i)} = \prod x_i y_i \cdot \prod a_{i\sigma(i)} = \prod a_{i\sigma(i)}$$

- They generate all other invariants
  - If $A$ not in null cone then $p(A) \neq 0$ for some matching
  - $A$ integer coeffs. & $p(A) \neq 0 \Rightarrow |p(A)|^2 \geq 1$

- $B \in G \cdot A \Rightarrow |p(B)|^2 = |p(A)|^2 \geq 1$

- $p(B)$ is a matching monomial $\Rightarrow ||B||^{2n} \geq |p(B)|^2 \geq 1$
Algorithm S – Analysis

Algorithm S: matrix $A$ integer entries bounded by $\nu$, param. $\epsilon > 0$.

Repeat $k$ times:
1. Normalize rows of $A$
2. Normalize columns of $A$

If at any point $db(A) \leq \epsilon$, output the scaling so far.
Else, output: no scaling.

Analysis [\sim LSW’00]:

1. $||A||^2 \leq \nu^2 \cdot n^2$ (bound on input)
2. $db(A) \geq \epsilon \Rightarrow ||A||^2$ decreases by $\exp(O(\epsilon))$
   after normalization (AM-GM)
3. $||B|| \geq 1$ for any scaling of $A$
Invariants – Operator Scaling

**Operator Scaling:** \( SL_n \times SL_n \sim M_n(\mathbb{C})^m \)

- Invariants: given \( B_i = LA_i R \)
  \[
  \det(\sum B_i \otimes Y_i) = \det(\sum (LA_i R) \otimes Y_i) \\
  = \det(\sum A_i \otimes Y_i) \cdot \det(L)^d \det(R)^d = \det(\sum A_i \otimes Y_i)
  \]
- They generate all other invariants
  - If \( (A_i) \) not in null cone then \( p(A) \neq 0 \) for some such inv.
- \( A_i, Y_i \) integer coeffs. & \( p(A) \neq 0 \) \( \Rightarrow |p(A)|^2 \geq 1 \)
  - \( B \in G \cdot A \Rightarrow |p(B)|^2 = |p(A)|^2 \geq 1 \)
    \[
    \Rightarrow \exp(nd) \cdot ||B||^{2nd} \geq |p(B)|^2 \geq 1 \\
    ||B|| \geq \exp(-n)
    \]
Algorithm G – Analysis

**Algorithm G**: tuple \((A_i)\) integer entries bounded by \(\nu, \epsilon > 0\).

Repeat \(k\) times:

1. Left normalize \((A_i) \rightarrow \sum A_i A_i^\dagger \sim I_n\)
2. Right normalize \((A_i) \rightarrow \sum A_i^\dagger A_i \sim I_n\)

If at any point \(dB(T) < \epsilon\), output scaling.
Else, output: **no scaling**.

**Analysis [GGOW’15]**:

1. \(\sum \|A_i\|^2 \leq \nu^2 \cdot n^2\) (bound on input)
2. \(dB(A) \geq \epsilon \Rightarrow \sum \|A_i\|^2\) decreases by \(\exp(O(\epsilon))\)
   after normalization (AM-GM)
3. \(\sum \|B_i\|^2 \geq \exp(-n)\) for any scaling of \(A\)
(Recap) Hilbert’s Foundational Results

Given vector space \( V \) and group \( G \) acting (linearly) on it

**Null cone** \( \mathcal{N}_G(V) = \{ v \in V \mid 0 \in G \cdot v \} \)

[Hil’93] Given vector space \( V \) and group \( G \) acting (linearly) on it \( \mathcal{N}_G(V) \) is the common zero set of all invariant polynomials. I.e.

\[
v \in \mathcal{N}_G(V) \iff p(v) = 0 \quad \forall \ p \text{ invariant}
\]

**Null-cone Problem:** given \( v \in V \), is \( v \in \mathcal{N}_G(V) \)?

Two ways of solving this problem!

- Optimization: \( \inf_{g \in G} (||g \cdot v||^2) \)
- Algebraic: decide if all invariants vanish (“PIT”)

Why are we talking about this? Where is DB?
Kempf-Ness & Non-commutative duality

Null-cone Problem: given \( \nu \in V \), is \( \nu \in \mathcal{N}_G(V) \) (i.e. \( 0 \in G \cdot \nu \))?

- Optimization: \( \text{cap}(\nu) = \inf_{g \in G} (||g \cdot \nu||^2) \)

How do we know we are “close” to the optimum?

- \[KN’79\] “Gradient is close to zero!”
  - Gradient “along the group action” (Lie Algebra)
  - General notion of convexity (geodesic-convexity)

\[KN’79\] “Non-commutative duality”

- \( \mu(\omega) \) moment map: gradient along group action (Ankit’s talk)
- Dual program: \( \text{cap}_\mu(\nu) = \inf_{g \in G} ||\mu(g \cdot \nu)||^2 \)

Far from DB \( \text{cap}_\mu(\nu) > 0 \iff \text{cap}(\nu) = 0 \)

In Null cone

\( \text{db}(A), \text{db}(T) \) norms of moment map for matrix/operator scaling!
Algorithm S – Primal dual approach

**Algorithm S**: matrix $A$ integer entries bounded by $v$, param. $\epsilon > 0$. 
Repeat $k$ times:
1. Normalize rows of $A$
2. Normalize columns of $A$
If at any point $db(A) \leq \epsilon$, output the scaling so far.
Else, output: no scaling.

**Analysis [~LSW’00]:**
1. $||A||^2 \leq v^2 \cdot n^2$ (bound on input)
2. $db(A) \geq \epsilon \Rightarrow ||A||^2$ decreases by $\exp(O(\epsilon))$ after normalization (AM-GM)
3. $||B|| \geq 1$ for any scaling of $A$
Invariant Theory – Orbit Closure Intersection

Invariant Theory:
\[ G = \mathbb{SL}_n(\mathbb{C})^2, \] vector space \[ V = M_n(\mathbb{C})^m \] action by L-R mult:
\[ (A_1, ..., A_m) \rightarrow (LA_1 R, ..., LA_m R) \]

Orbit Closure: given \[ v = (A_1, ..., A_m) \in V, \] orbit closure is
\[ \overline{O_v} = \{(LA_1 R, ..., LA_m R) \mid (L, R) \in G\} \]

Orbit Closure Intersection Problem: given two quantum operators \[ u = (A_1, ..., A_m), \ v = (B_1, ..., B_m), \] is \[ \overline{O_u} \cap \overline{O_v} \neq \emptyset \]?

If \[ v = 0 \] problem becomes the null-cone problem. From [GGOW’16]: connections to non-commutative PIT, non-commutative algebra, combinatorics, functional analysis...

How can we solve the orbit intersection problem for L-R action?
What do we need to do?

Why is Operator Balancing not enough?

- Orbit closures can be exponentially close and not intersect
  - Need to have $\epsilon = \exp(-\text{poly}(n))$ approximation
  - Not the case for null-cone problem
- Operator Balancing runs in time $\text{poly}(n/\epsilon)$
  - Only good for null cone

We need $\log(1/\epsilon)$ convergence!

How to get it? Different algorithm!
KN’79 – Duality Theory

[KN’79]:

- Elts of min norm in $\mathcal{O}_{(A_1, \ldots, A_m)}$, are DB operators
  - $\epsilon$-close to DB implies $\epsilon$-close to min. norm
- $(B_1, \ldots, B_m)$ and $(C_1, \ldots, C_m)$ of minimum norm in $\mathcal{O}_{(A_1, \ldots, A_m)}$ then equivalent under unitary

[AGLOW’18]: solving orbit closure intersection problem

1. g-convex opt finds $\epsilon$-approx to element of minimum norm (DB)
2. With elements of min norm, test if they are $SU(n)$-equivalent

What is this g-convexity?
Convexity

Convexity (Euclidean geometry):

• Shortest path between \( A, B \) given by line

• Convex Set \( \mathcal{K} \):

• Convex function:

Convex

\( g(\cdot) \)

Ellipsoid

Interior Point Methods

2\textsuperscript{nd} order methods
What is Geodesic Convexity?

**Geodesic Convexity:**

- Shortest path between $A, B$ given by geodesic
- Geodesically Convex Set $\mathcal{K}$: $A, B \in \mathcal{K}$ so is its geodesics
- Geodesically Convex function: $f$ convex along each geodesic!
Geodesic Convexity

Example (our setup): complex positive definite matrices $\mathcal{S}_+$ with geodesic from $A$ to $B$ given by:

$$\gamma_{A,B} : [0, 1] \rightarrow \mathcal{S}_+ \quad \gamma_{A,B}(t) = A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2}$$

Convexity:

- $K \subseteq \mathcal{S}_+$ is g-convex if $\forall A, B \in K$ geodesic from $A$ to $B$ in $K$
- Function $f : K \rightarrow \mathbb{R}$ is g-convex if univariate function $f(\gamma_{A,B}(t))$ is convex in $t$ for any $A, B \in K$
Geodesically Convex Functions

Geodesically convex functions over $S_+$:

- \( \log(||g \cdot v||^2) \)

- \( \log(g \cdot g^\dagger) \) (geodesically linear)

Log of capacity $\text{def} \ \log(||g \cdot v||^2) - \log(g \cdot g^\dagger)$ g-convex!

For $\log(1/\epsilon)$ convergence, need new opt. tools for g-convex fncts.

No analog of *ellipsoid* or *interior point method* known for this setting.
Self Concordance & Self Robustness

**Self concordance:** $f : \mathbb{R} \to \mathbb{R}$ is self concordant if

$$|f'''(x)| \leq 2(f''(x))^{3/2}$$

$f : \mathbb{R}^n \to \mathbb{R}$ self concordant if self concordant along each line.

$h : S_+ \to \mathbb{R}$ g-self concordant if self concordant along each geodesic.

Unfortunately, log of capacity **NOT** self-concordant.

**Self robustness:** $f : \mathbb{R} \to \mathbb{R}$ is self robust if

$$|f'''(x)| \leq 2 \cdot f''(x)$$

$f : \mathbb{R}^n \to \mathbb{R}$ self robust if self robust along each line.

$h : S_+ \to \mathbb{R}$ g-self robust if self robust along each geodesic.

Log of capacity is geodesically self-robust!

**Question:** Can we efficiently optimize g-self robust functions?
This work – g-convex opt for self-robust fcns

Problem: given \( f : S_+ \to \mathbb{R} \) g-self robust, \( \epsilon > 0 \), and bound on initial distance \( R \) to OPT (diameter) find \( X_\epsilon \in S_+ \) such that

\[
f(X_\epsilon) \leq \inf_{Y \in S_+} f(Y) + \epsilon
\]

Theorem [AGLOW’18]:
There exists a deterministic \( \text{poly}(n, R, \log(1/\epsilon)) \), algorithm for the problem above.

- Second order method, generalizing recent work of [ALOW’17, CMTV’17] for matrix scaling to g-convex setting
- Generalizes to other manifolds and metrics

Remark:
- For operator scaling, \( X_\epsilon \) also gives us scaling \( \epsilon \)-close to DB
Problem: given \( f : S_+ \to \mathbb{R} \) g-self robust, \( \epsilon > 0 \), and bound on initial distance \( R \) to OPT (diameter) find \( X_\epsilon \in S_+ \) such that
\[
f(X_\epsilon) \leq \inf_{Y \in S_+} f(Y) + \epsilon
\]

Algorithm

- Start with \( X_0 = I \), \( \ell = O(R \cdot \log(1/\epsilon)) \).
- For \( t = 0 \) to \( \ell - 1 \)
  - \( f(t)(D) = f(X_t^{1/2} \exp(D)X_t^{1/2}) \).
  - \( Q_t \) quadratic-approximation to \( f(t) \).
  - \( D_t = \text{argmin}_{\|D\|_F \leq 1} Q_t(D) \). \( (\text{Euclidean convex opt.}) \)
  - \( X_{t+1} = X_t^{1/2} \exp(D_t)X_t^{1/2} \).
- Return \( X_\ell \).

• Why would we need this instead of regular scaling?
• What is the bound for \( R \) in operator scaling?
  • \([\text{AGLOW'18}]\) polynomial bound for \( R \)
Why do we need $\log(1/\epsilon)$ convergence?

- Orbit closures can be exponentially close and not intersect
  - Need to have $\epsilon = \exp(-\text{poly}(n))$ approximation
- Not the case for null-cone problem
- $SU(n)$-equivalence algorithm also approximate (and lossy)

[AGLOW'18]: solving orbit closure intersection problem

1. g-convex opt finds $\epsilon$-approx to element of minimum norm (DB)
2. With elements of min norm, test if they are $SU(n)$-equivalent
Advertisement

Amazing workshop at the IAS!
Videos & materials online
https://www.math.ias.edu/ocit2018

Survey on all of this (w/ Ankit) on arxiv & on EATCS complexity column!
(link on my webpage)
Open Questions

• Complexity of null-cone problem? Of OCI?

• Better algorithms for scaling problems?
  • Best algorithms we have are $\text{poly}(R \cdot \log(1/\epsilon))$

• Efficient algorithms for null-cone and orbit closure intersection for more general actions?
  • Recent developments for general scaling, though still $\text{poly}(n/\epsilon)$
  • Upcoming work gets $\text{poly}(R\log(1/\epsilon))$, but still have bad bounds on $R$

Thank you!