Operator scaling & doubly-stochastic quantum channels

Rafael Oliveira
University of Waterloo

Fields Institute
July 10, 2025

Quantum Channels

- Quantum channels: completely positive, trace preserving
- Choi's theorem: completely positive operators given by Kraus operators $A := (A_1, \ldots, A_m) \in (\mathbb{C}^{n\times n})^m$

$$\Phi_A(X) := \sum_{i=1}^{m} A_i X A_i^*, \quad \Phi_A^*(X) = \sum_{i=1}^{m} A_i^* X A_i$$

Trace preserving: $\Phi_A^*(I) = I$.

Quantum Channels

- Quantum channels: completely positive, trace preserving
- Choi's theorem: completely positive operators given by Kraus operators $A := (A_1, \ldots, A_m) \in (\mathbb{C}^{n\times n})^m$

$$\Phi_A(X) := \sum_{i=1}^{m} A_i X A_i^*, \quad \Phi_A^*(X) = \sum_{i=1}^{m} A_i^* X A_i$$

Trace preserving: $\Phi_A^*(I) = I$.

- Norm of an operator:

$$\|A\| := \left(\mathrm{Tr}\left[\sum_{i=1}^{m} A_i^* A_i\right]\right)^{1/2} = (\mathrm{Tr}[\Phi_A^*(I)])^{1/2} = (\mathrm{Tr}[\Phi_A(I)])^{1/2}$$

Quantum Channels

- Quantum channels: completely positive, trace preserving
- Choi's theorem: completely positive operators given by Kraus operators $A := (A_1, \ldots, A_m) \in (\mathbb{C}^{n \times n})^m$

$$\Phi_A(X) := \sum_{i=1}^{m} A_i X A_i^*, \quad \Phi_A^*(X) = \sum_{i=1}^{m} A_i^* X A_i$$

Trace preserving: $\Phi_A^*(I) = I$.

- Norm of an operator:

$$\|A\| := \left(\mathrm{Tr} \left[\sum_{i=1}^{m} A_i^* A_i \right] \right)^{1/2} = (\mathrm{Tr}[\Phi_A^*(I)])^{1/2} = (\mathrm{Tr}[\Phi_A(I)])^{1/2}$$

- Transport plan from $\rho \in \mathrm{PSD}(n)$ to $\sigma \in \mathrm{PSD}(n)$:

$$\Phi_A(\rho) = \sigma \quad \text{and} \quad \Phi_A^*(I) = I$$

Alternatively: setting $B_i := A_i \rho^{1/2}$, we get

$$\Phi_B(I) = \sigma \quad \text{and} \quad \Phi_B^*(I) = \rho$$

Operator Scaling - Setup

- **Input:** Kraus operators $A := (A_1, \ldots, A_m) \in (\mathbb{C}^{n \times n})^m$ of our quantum channel Φ_A.

Operator Scaling - Setup

- **Input:** Kraus operators $A := (A_1, \ldots, A_m) \in (\mathbb{C}^{n\times n})^m$ of our quantum channel Φ_A.
- Scaling of a channel: $L, R \in SL(n)$

$$(L, R) \circ A \mapsto LAR^T := (LA_1R^T, \ldots, LA_mR^T)$$

Operator Scaling - Setup

- **Input:** Kraus operators $A := (A_1, \ldots, A_m) \in (\mathbb{C}^{n \times n})^m$ of our quantum channel Φ_A.
- Scaling of a channel: $L, R \in SL(n)$

$$(L, R) \circ A \mapsto LAR^T := (LA_1R^T, \ldots, LA_mR^T)$$

- **Output:** are there scalings $(L, R) \in SL(n)^2$ such that the operator $B := LAR^T$ is doubly-stochastic?

$$\Phi_B(I) = \Phi_B^*(I) = \frac{\|B\|^2}{n} \cdot I$$

  - Important: enough to scale over $\mathrm{PD}(n)^2$
  - Polar decomposition (unitaries cancel)

Operator Scaling - Setup

- **Input:** Kraus operators $A := (A_1, \ldots, A_m) \in (\mathbb{C}^{n\times n})^m$ of our quantum channel Φ_A.
- Scaling of a channel: $L, R \in SL(n)$

$$(L, R) \circ A \mapsto LAR^T := (LA_1R^T, \ldots, LA_mR^T)$$

- **Output:** are there scalings $(L, R) \in SL(n)^2$ such that the operator $B := LAR^T$ is doubly-stochastic?

$$\Phi_B(I) = \Phi_B^*(I) = \frac{\|B\|^2}{n} \cdot I$$

  - Important: enough to scale over $\mathrm{PD}(n)^2$
  - Polar decomposition (unitaries cancel)
- Can also scale to arbitrary marginals: (see references)

$$\Phi_B(I) = \rho, \quad \Phi_B^*(I) = \sigma$$

Operator Scaling Applications

- Functional Analysis: computation of Brascamp-Lieb constants
- Schrödinger bridge problem
- Schur-Horn problem
- Paulsen problem
- Non-commutative algebra: word problem for free skew field
- Extremal combinatorics: more quantitative generalizations of Sylvester-Gallai theorems
- sample complexity & algorithm for matrix normal model
- For more, see **[Idel 2016]** and **[Garg O. 2018]**.

Operator Scaling - Gurvits Algorithm

1. **Input:** Kraus operators $A \in (\mathbb{Z}[i]^{n \times n})^m$, approximation parameter $\varepsilon > 0$.

Operator Scaling - Gurvits Algorithm

1. **Input:** Kraus operators $A \in (\mathbb{Z}[i]^{n\times n})^m$, approximation parameter $\varepsilon > 0$.
2. Let

$$\Delta(A)^2 := \left\| \frac{1}{\|A\|^2} \cdot \Phi_A(I) - \frac{1}{n} I_n \right\|^2 + \left\| \frac{1}{\|A\|^2} \cdot \Phi_A^*(I) - \frac{1}{n} I_n \right\|^2$$

Operator Scaling - Gurvits Algorithm

1. **Input:** Kraus operators $A \in (\mathbb{Z}[i]^{n\times n})^m$, approximation parameter $\varepsilon > 0$.
2. Let

$$\Delta(A)^2 := \left\| \frac{1}{\|A\|^2} \cdot \Phi_A(I) - \frac{1}{n} I_n \right\|^2 + \left\| \frac{1}{\|A\|^2} \cdot \Phi_A^*(I) - \frac{1}{n} I_n \right\|^2$$

3. For T iterations, alternate between:
   - Left normalize A, i.e.:

$$A^{(t+1)} := \left(\Phi_{A^{(t)}}(I)\right)^{-1/2} \cdot A^{(t)}$$

   - Right normalize A, i.e.:

$$A^{(t+1)} := A^{(t)} \cdot \left(\Phi^*_{A^{(t)}}(I)\right)^{-1/2}$$

Operator Scaling - Gurvits Algorithm

1. **Input:** Kraus operators $A \in (\mathbb{Z}[i]^{n\times n})^m$, approximation parameter $\varepsilon > 0$.
2. Let
$$\Delta(A)^2 := \left\| \frac{1}{\|A\|^2} \cdot \Phi_A(I) - \frac{1}{n} I_n \right\|^2 + \left\| \frac{1}{\|A\|^2} \cdot \Phi_A^*(I) - \frac{1}{n} I_n \right\|^2$$
3. For T iterations, alternate between:
   - Left normalize A, i.e.:
   $$A^{(t+1)} := \left(\Phi_{A^{(t)}}(I)\right)^{-1/2} \cdot A^{(t)}$$
   - Right normalize A, i.e.:
   $$A^{(t+1)} := A^{(t)} \cdot \left(\Phi^*_{A^{(t)}}(I)\right)^{-1/2}$$
4. If at any point $\Delta(A^{(t)}) \leq \varepsilon$ return.

Operator Scaling (via Invariant Theory)

- **Function of interest:** log norm

Operator Scaling (via Invariant Theory)

- **Function of interest:** log norm

$$F_A : \mathrm{PD}(n) \times \mathrm{PD}(n) \to \mathbb{R}$$

$$\begin{aligned} F_A(P, Q) &:= \log(\|P^{1/2} A Q^{1/2}\|^2) - \frac{1}{n} \cdot (\log\det P + \log\det Q) \\ &= \log\left(\sum_{i=1}^{m} \mathrm{Tr}[P A_i Q A_i^*]\right) - \frac{1}{n} \cdot (\log\det P + \log\det Q) \end{aligned}$$

Operator Scaling (via Invariant Theory)

- **Function of interest:** log norm

$$F_A : \mathrm{PD}(n) \times \mathrm{PD}(n) \to \mathbb{R}$$

$$\begin{aligned} F_A(P, Q) &:= \log(\|P^{1/2} A Q^{1/2}\|^2) - \frac{1}{n} \cdot (\log\det P + \log\det Q) \\ &= \log\left(\sum_{i=1}^{m} \mathrm{Tr}[P A_i Q A_i^*]\right) - \frac{1}{n} \cdot (\log\det P + \log\det Q) \end{aligned}$$

- **Problem of interest:** **[Gurvits 2004]**

$$\mathrm{logcap}(A) := \inf_{(P,Q) \in \mathrm{PD}(n)^2} F_A(P, Q)$$

Geometry

Which geometry is the most appropriate?

- $\mathrm{PD}(n)$ is a Hadamard manifold[1]
  - complete, simply connected Riemannian manifold
  - non-positive sectional curvature

[1] Metric arising from Hessian of log-determinant.

Geometry

Which geometry is the most appropriate?

- $\mathrm{PD}(n)$ is a Hadamard manifold[1]
  - complete, simply connected Riemannian manifold
  - non-positive sectional curvature
- Tangent space at $\Theta \in \mathrm{PD}(n)$: Hermitian matrices

[1] Metric arising from Hessian of log-determinant.

Geometry

Which geometry is the most appropriate?

- $\mathrm{PD}(n)$ is a Hadamard manifold[1]
  - complete, simply connected Riemannian manifold
  - non-positive sectional curvature
- Tangent space at $\Theta \in \mathrm{PD}(n)$: Hermitian matrices
- Geodesics from Θ (in direction H) are

$$\gamma_{\Theta}(t) := \Theta^{1/2} e^{tH} \Theta^{1/2}$$

[1] Metric arising from Hessian of log-determinant.

Geometry

Which geometry is the most appropriate?

- $\mathrm{PD}(n)$ is a Hadamard manifold[1]
  - complete, simply connected Riemannian manifold
  - non-positive sectional curvature
- Tangent space at $\Theta \in \mathrm{PD}(n)$: Hermitian matrices
- Geodesics from Θ (in direction H) are

$$\gamma_\Theta(t) := \Theta^{1/2} e^{tH} \Theta^{1/2}$$

- $F_A(P, Q)$ is geodesically convex!

[1] Metric arising from Hessian of log-determinant.

Geometry

Which geometry is the most appropriate?

- $\mathrm{PD}(n)$ is a Hadamard manifold[1]
  - complete, simply connected Riemannian manifold
  - non-positive sectional curvature
- Tangent space at $\Theta \in \mathrm{PD}(n)$: Hermitian matrices
- Geodesics from Θ (in direction H) are

$$\gamma_\Theta(t) := \Theta^{1/2} e^{tH} \Theta^{1/2}$$

- $F_A(P, Q)$ is geodesically convex!
- $F_A(P, Q) = F_{P^{1/2} A Q^{1/2}}(I, I)$

[1] Metric arising from Hessian of log-determinant.

Operator Scaling

(Riemannian) Gradient of log-norm (at identity):

$$
\begin{aligned}
&\langle(\nabla F_A)(I,I),(R,S)\rangle = \partial_{t=0} F_A(e^{tR},e^{tS}) \\
&= \frac{1}{\|A\|^2}\cdot\left(\operatorname{Tr}[R\cdot\Phi_A(I)] + \operatorname{Tr}[S\Phi_A^*(I)]\right) - \frac{1}{n}\cdot\operatorname{Tr} R - \frac{1}{n}\cdot\operatorname{Tr} S \\
&= \left\langle \left(\frac{1}{\|A\|^2}\cdot\Phi_A(I) - \frac{1}{n}I, \frac{1}{\|A\|^2}\cdot\Phi_A^*(I) - \frac{1}{n}I\right), (R,S)\right\rangle
\end{aligned}
$$

Operator Scaling

(Riemannian) Gradient of log-norm (at identity):

$$\begin{aligned}
&\langle (\nabla F_A)(I,I),(R,S)\rangle = \partial_{t=0} F_A(e^{tR},e^{tS}) \\
&= \frac{1}{\|A\|^2}\cdot(\operatorname{Tr}[R\cdot\Phi_A(I)]+\operatorname{Tr}[S\Phi_A^*(I)]) - \frac{1}{n}\cdot\operatorname{Tr} R - \frac{1}{n}\cdot\operatorname{Tr} S \\
&= \left\langle \left(\frac{1}{\|A\|^2}\cdot\Phi_A(I)-\frac{1}{n}I, \frac{1}{\|A\|^2}\cdot\Phi_A^*(I)-\frac{1}{n}I\right),(R,S)\right\rangle
\end{aligned}$$

Which implies:

$$(\nabla F_A)(I,I) = \left(\frac{1}{\|A\|^2}\cdot\Phi_A(I)-\frac{1}{n}I, \frac{1}{\|A\|^2}\cdot\Phi_A^*(I)-\frac{1}{n}I\right)$$

Operator Scaling

(Riemannian) Gradient of log-norm (at identity):

$$\begin{aligned}
\langle (\nabla F_A)(I,I),(R,S)\rangle &= \partial_{t=0} F_A(e^{tR},e^{tS}) \\
&= \frac{1}{\|A\|^2}\cdot\left(\mathrm{Tr}[R\cdot\Phi_A(I)]+\mathrm{Tr}[S\Phi_A^*(I)]\right)-\frac{1}{n}\cdot\mathrm{Tr}\,R-\frac{1}{n}\cdot\mathrm{Tr}\,S \\
&= \left\langle\left(\frac{1}{\|A\|^2}\cdot\Phi_A(I)-\frac{1}{n}I,\frac{1}{\|A\|^2}\cdot\Phi_A^*(I)-\frac{1}{n}I\right),(R,S)\right\rangle
\end{aligned}$$

Which implies:

$$(\nabla F_A)(I,I)=\left(\frac{1}{\|A\|^2}\cdot\Phi_A(I)-\frac{1}{n}I,\frac{1}{\|A\|^2}\cdot\Phi_A^*(I)-\frac{1}{n}I\right)$$

- norm of gradient is how far we are from doubly-stochastic!

Operator Scaling - Analysis

Same approach as in matrix scaling.

- Polynomial invariants to lower bound the minimum log-norm
  - analyze alternating minimization as in matrix scaling
- If all invariants vanish, then log-norm unbounded (not scalable to doubly stochastic)
  - quantify how far from doubly-stochastic we will be

Invariants

Our setup: $G = SL(n) \times SL(n)$ acting on $V = (\mathbb{C}^{n\times n})^m$

$$(L,R) \circ A := (A_1, \dots, A_m) \mapsto LAR^T := (LA_1R^T, \dots, LA_mR^T)$$

Theorem (Invariants for LR-action)

The (homogeneous) invariants for above action are of the form

$$h_M(Z_1, \dots, Z_m) := \det\left(\sum_{i=1}^{m} M_i \otimes Z_i\right)$$

where $M \in (\mathbb{C}^{d\times d})^m$.

Theorem (Degree bounds)

Invariants are generated as a $\mathbb{C}$*-algebra by the above polynomials with degree* $d \leq n^6 m$

Invariants and Capacity lower bound

Lemma

If there is $M \in (\mathbb{C}^{d\times d})^m$ such that $h_M(A) \neq 0$, then there is $M \in ([2nd]^{d\times d})^m$.
Moreover, for such M, the coefficients of $h_M(Z)$ are bounded by $\exp(O(dn \cdot \log(dn)))$.

- Moreover part: note that each coefficient of h_M is upper bounded by

$$\mathrm{Per}\left(\sum_{i=1}^{m} M_i \otimes J\right) \leq \prod_{i=1}^{dn}(\text{row sums})$$

Corollary

If $A \in (\mathbb{Z}[i]^{n\times n})^m$ doesn't vanish on some invariant, then

$$\mathrm{logcap}(A) \geq \frac{1}{\mathrm{poly}(n)}$$

Invariants and 1-PSGs

Theorem (Hilbert-Mumford)

If all invariants vanish on A, there is a 1-PSG $\lambda : \mathbb{C}^* \to G$ s.t.

$$\lim_{t\to\infty} \lambda(t) \circ A = 0.$$

- 1-PSGs of G are parameterized by $(L, R) \in G$ and $(\vec{a}, \vec{b}) \in \mathbb{Z}^n \times \mathbb{Z}^n$ with $\vec{a}, \vec{b} \perp \vec{1}$.

$$\lambda_{(L,R),(\vec{a},\vec{b})}(t) := \left(L^{-1}\,\mathrm{diag}(t^{a_1}, \cdots, t^{a_n})L, R^{-1}\,\mathrm{diag}(t^{b_1}, \cdots, t^{b_n})R\right)$$

Invariants and 1-PSGs

Theorem (Hilbert-Mumford)

If all invariants vanish on A, there is a 1-PSG $\lambda : \mathbb{C}^* \to G$ s.t.

$$\lim_{t\to\infty} \lambda(t) \circ A = 0.$$

- 1-PSGs of G are parameterized by $(L, R) \in G$ and $(\vec{a}, \vec{b}) \in \mathbb{Z}^n \times \mathbb{Z}^n$ with $\vec{a}, \vec{b} \perp \vec{1}$.

$$\lambda_{(L,R),(\vec{a},\vec{b})}(t) := \Big(L^{-1}\operatorname{diag}(t^{a_1}, \cdots, t^{a_n})L, R^{-1}\operatorname{diag}(t^{b_1}, \cdots, t^{b_n})R\Big)$$

- How can it send A to 0?

$$\lambda(t) \circ A = (L^{-1}, R^{-1}) \circ (\operatorname{diag}(t^{\vec{a}}), \operatorname{diag}(t^{\vec{b}})) \circ (L, R) \circ A$$

Invariants and 1-PSGs

Theorem (Hilbert-Mumford)

If all invariants vanish on A, there is a 1-PSG $\lambda : \mathbb{C}^* \to G$ s.t.

$$\lim_{t\to\infty} \lambda(t) \circ A = 0.$$

- 1-PSGs of G are parameterized by $(L, R) \in G$ and $(\vec{a}, \vec{b}) \in \mathbb{Z}^n \times \mathbb{Z}^n$ with $\vec{a}, \vec{b} \perp \vec{1}$.

$$\lambda_{(L,R),(\vec{a},\vec{b})}(t) := \left(L^{-1} \operatorname{diag}(t^{a_1}, \cdots, t^{a_n})L, R^{-1} \operatorname{diag}(t^{b_1}, \cdots, t^{b_n})R\right)$$

- How can it send A to 0?

$$\lambda(t) \circ A = (L^{-1}, R^{-1}) \circ (\operatorname{diag}(t^{\vec{a}}), \operatorname{diag}(t^{\vec{b}})) \circ (L, R) \circ A$$

- Only happens if $\lim_{t\to\infty}(\operatorname{diag}(t^{\vec{a}}), \operatorname{diag}(t^{\vec{b}})) \circ B = 0$ where $B = (L, R) \circ A$.

Invariants and 1-PSGs

Theorem (Hilbert-Mumford)

If all invariants vanish on A, there is a 1-PSG $\lambda : \mathbb{C}^* \to G$ s.t.

$$\lim_{t\to\infty} \lambda(t) \circ A = 0.$$

- 1-PSGs of G are parameterized by $(L, R) \in G$ and $(\vec{a}, \vec{b}) \in \mathbb{Z}^n \times \mathbb{Z}^n$ with $\vec{a}, \vec{b} \perp \vec{1}$.

$$\lambda_{(L,R),(\vec{a},\vec{b})}(t) := \left(L^{-1}\operatorname{diag}(t^{a_1}, \cdots, t^{a_n})L, R^{-1}\operatorname{diag}(t^{b_1}, \cdots, t^{b_n})R\right)$$

- How can it send A to 0?

$$\lambda(t) \circ A = (L^{-1}, R^{-1}) \circ (\operatorname{diag}(t^{\vec{a}}), \operatorname{diag}(t^{\vec{b}})) \circ (L, R) \circ A$$

- Only happens if $\lim_{t\to\infty}(\operatorname{diag}(t^{\vec{a}}), \operatorname{diag}(t^{\vec{b}})) \circ B = 0$ where $B = (L, R) \circ A$.
- Same condition as we saw in matrix scaling! (Hall blocker)

Alternating Minimization - Analysis

Lemma (Bound on distance to doubly stochastic)

If all invariants vanish on A, then $\Delta(A) \geq 1/n^2$.

Lemma (Quantitative AM-GM)

Let $Z \in \mathrm{PD}(n)$ with $\operatorname{Tr} Z = n$. Then

$$\log\det Z \leq -\frac{1}{6} \cdot \min\{\|Z - I\|_F^2, 1\}.$$

- Setting $\varepsilon = 1/2n^2$ and running the algorithm, we have:
  - $f(\tau) := \log \left\| A^{(\tau)} \right\|$ implies

    $$f(\tau) - f(\tau - 1) = \frac{1}{n} \cdot \left(\log\det \Phi_{A^{(\tau-1)}}(I) + \log\det \Phi^*_{A^{(\tau-1)}}(I)\right)$$

  - $\Delta(A^{(\tau-1)}) \geq 1/2n^2$ quantitative AM-GM lemma implies

    $$f(\tau) - f(\tau - 1) \leq -\frac{1}{24n^4}$$

References I

Ayers, P. and Coudreuse, F. and Gerolin, A. and Oliveira, R. and Ramachandran, A. (2025)
Operator Scaling & Quantum Optimal Transport
Manuscript

Berta, F. and Fawzi, O. and Tomamichel, M. (2017)
On Variational Expressions for Quantum Relative Entropies
Letters in Mathematical Physics, 107, pp.2239-2265.

Gurvits, L. (2004)
Classical complexity and quantum entanglement.
Journal of Computer and System Sciences, 69(3), pp.448-484.

Idel, M. (2016)
A review of matrix scaling and Sinkhorn's normal form for matrices and positive maps
arXiv preprint arXiv:1609.06349

References II

- Garg, A. and Gurvits, L. and Oliveira, R. and Wigderson, A. (2018)
  Operator Scaling: Theory and Applications
  Foundations of Computational Mathematics
- Garg, A. and Oliveira, R. (2018)
  Recent progress on scaling algorithms and applications
  Bulletin of EATCS 2.125
- Peyré, G. and Cuturi, M. (2019)
  Computational optimal transport: With applications to data science.
  Foundations and Trends in Machine Learning, 11(5-6), pp.355-607.
- Sinkhorn, R. (1964)
  A relationship between arbitrary positive matrices and doubly stochastic matrices.
  The annals of mathematical statistics, 35(2), pp.876-879.

References III

Silva Louzeiro, M., Bergmann, R. and Herzog, R. (2022)
Fenchel duality and a separation theorem on Hadamard manifolds.
SIAM Journal on Optimization, 32(2), pp.854-873.

Keyl, M., 2006.
Quantum state estimation and large deviations.
Reviews in Mathematical Physics, 18(01), pp.19-60.

Keyl, M. and Werner, R.F., 2001.
Estimating the spectrum of a density operator.
Physical Review A, 64(5), p.052311.