Rafael Oliveira
University of Toronto

Analysis of Scaling Algorithms for Matrix & Operator Scaling

1 Contents

- Scaling Algorithms
- Three Step Analysis
- Generalization

One More Application of Scaling

Non-Negative Matrices & Scaling

$X \in M_n(\mathbb{R}_{\geq 0})$ is **doubly stochastic (DS)** if row/column sums of X are equal to **1**.

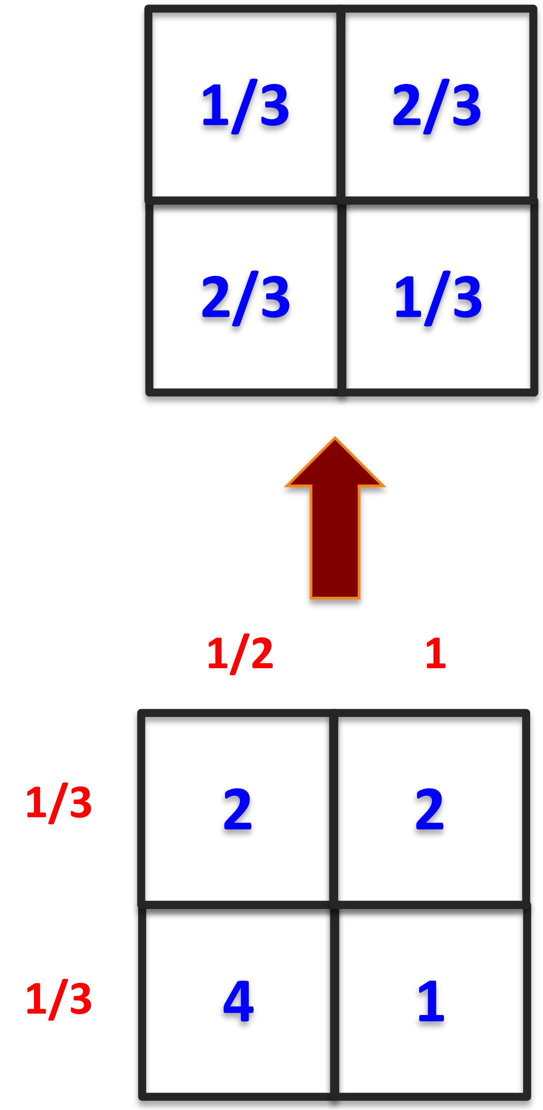
Y is **scaling** of X if ∃ positive $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$ s.t. $y_{ij} = \alpha_i x_{ij} \beta_j$.

$\boldsymbol{X}$ has DS scaling if ∃ scaling $\mathbf{Y}$ of $\boldsymbol{X}$ s.t. all row/column sums of $\boldsymbol{Y}$ equal **1**.

$$\boldsymbol{ds}(\boldsymbol{A}) = \sum_{\boldsymbol{i}} (\boldsymbol{r}_{\boldsymbol{i}} - \mathbf{1})^2 + \sum_{\boldsymbol{j}} (\boldsymbol{c}_{\boldsymbol{j}} - \mathbf{1})^2$$

$\boldsymbol{A}$ has approx. DS scaling if $\forall \boldsymbol{\epsilon} > \mathbf{0}$ there is scaling $\boldsymbol{B}_{\boldsymbol{\epsilon}}$ of $\boldsymbol{A}$ s.t. $\mathbf{ds}(\mathbf{B}_{\boldsymbol{\epsilon}}) < \boldsymbol{\epsilon}$.

1. When does $\boldsymbol{X}$ have approx. DS scaling?
2. Can we find it efficiently?



Analysis (Ankit's talk)

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

Repeat $\boldsymbol{k}$ times:

1. Normalize rows of $\boldsymbol{A}$ (make $\boldsymbol{r_i} = \mathbf{1}$)
2. Normalize columns of $\boldsymbol{A}$ (make $\boldsymbol{c_j} = \mathbf{1}$)

If at any point $\mathbf{ds}(\boldsymbol{A}) < \boldsymbol{\epsilon}$, output the scaling so far.

Else, output: **no scaling.**

Analysis [LSW'00]:

1. $\boldsymbol{Per}(\boldsymbol{A}) > \mathbf{0} \Rightarrow \boldsymbol{Per}(\boldsymbol{A}) > \boldsymbol{v}^{-\boldsymbol{n}}$
2. $\mathbf{ds}(\mathbf{A}) \geq \boldsymbol{\epsilon} \Rightarrow \boldsymbol{Per}(\boldsymbol{A})$ grows by $\boldsymbol{exp}(\boldsymbol{O}(\boldsymbol{\epsilon}))$ after each normalization
3. $\boldsymbol{Per}(\boldsymbol{A}) \leq \mathbf{1}$ for any normalized matrix

Within $\boldsymbol{poly}(\boldsymbol{n}/\boldsymbol{\epsilon})$ iterations we will get our scaling!

Quantum Operators – Recap of Definition

Quantum operator: $\mathbf{T}\colon \mathbf{M}_n(\mathbb{C}) \to M_n(\mathbb{C})$ given by $(A_1, \dots, A_m)$ s.t.

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

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

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

$\mathbf{T}\colon \mathbf{M}_n(\mathbb{C}) \to M_n(\mathbb{C})$ is **doubly stochastic** if $T(I) = T^*(I) = I$.

Distance to doubly-stochastic:

$$ds(T) \stackrel{\text{def}}{=} \|T(I) - I\|_F^2 + \|T^*(I) - I\|_F^2$$

Scaling $T_{L,R}(X)$ of $T(X)$ consists of $L, R \in SL(n)$ s.t.

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

Operator Scaling – Algorithm G

Problem: operator $\mathbf{T} = (A_1, \dots, A_m)$, $\epsilon > 0$, can T be ϵ-scaled to double stochastic? If yes, find scaling.

Algorithm G [Gurvits' 04]:

Repeat k times:

1. Left normalize: $(A_1, \dots, A_m) \leftarrow (LA_1, \dots, LA_m)$ s.t. $T(I) = I$
2. Right normalize: $(A_1, \dots, A_m) \leftarrow (A_1R, \dots, A_mR)$ s.t. $T^*(I) = I$

If at any point $\mathbf{ds}(\mathbf{T}) < \epsilon$ output scaling.

Else output **no scaling**.

Potential Function (Capacity) [Gurvits'04]:

$$cap(T) = inf \left\{ \frac{det(T(X))}{det(X)} : X \succ 0 \right\}.$$

For $\epsilon < 1/n^2$, can scale T to ϵ-close to DS iff $cap(T) > 0$.

Algorithm G – Analysis

Algorithm G:

Repeat $\boldsymbol{k}$ times:

1. Left normalize: $(\boldsymbol{A_1}, \dots, \boldsymbol{A_m}) \leftarrow (\boldsymbol{LA_1}, \dots, \boldsymbol{LA_m})$ s.t. $\boldsymbol{T}(\boldsymbol{I}) = \boldsymbol{I}$.
2. Right normalize: $(\boldsymbol{A_1}, \dots, \boldsymbol{A_m}) \leftarrow (\boldsymbol{A_1 R}, \dots, \boldsymbol{A_m R})$ s.t. $\boldsymbol{T}^*(\boldsymbol{I}) = \boldsymbol{I}$.

If at any point $\mathbf{ds}(\boldsymbol{T}) < \boldsymbol{\epsilon}$, output current scaling.

Else output **no scaling**.

Potential Function (Capacity) [Gur'04]:

$$\boldsymbol{cap}(\boldsymbol{T}) = \boldsymbol{inf}\left\{\frac{\boldsymbol{det}(\boldsymbol{T}(\boldsymbol{X}))}{\boldsymbol{det}(\boldsymbol{X})} : \boldsymbol{X} \succ \mathbf{0}\right\}.$$

Analysis [Gur'04, GGOW'15]:

1. $\boldsymbol{cap}(\boldsymbol{T}) > \mathbf{0} \Rightarrow \boldsymbol{cap}(\boldsymbol{T}) > \boldsymbol{e}^{-\boldsymbol{poly}(\boldsymbol{n})}$ (GGOW'15)
2. $\boldsymbol{ds}(\boldsymbol{T}) \geq \boldsymbol{\epsilon} \Rightarrow \boldsymbol{cap}(\boldsymbol{T})$ grows by $\times \boldsymbol{exp}(\boldsymbol{O}(\boldsymbol{\epsilon}))$ after normalization
3. $\mathbf{cap}(\mathbf{T}) \leq \mathbf{1}$ for normalized operators.

Analysis – Step 2

Lemma [LSW'00]: $a_1, \dots, a_n > 0$ s.t. $\sum a_i = n$ and $\sum (a_i - 1)^2 = \Delta$

$$\prod a_i \le exp(-\Delta/6)$$

Claim: assume $T^*(I) = I$.

$$ds(T) > \epsilon \Rightarrow det\big(T(I)\big) \le exp(-\epsilon/6)$$

Proof sketch:

$$tr\big(T(I)\big) = tr\big(\sum A_i A_i^\dagger\big) = tr\big(\sum A_i^\dagger A_i\big) = tr\big(T^*(I)\big) = n$$

$$T(I) = \sum \lambda_i u_i u_i^\dagger \text{ where } \lambda_i > 0$$

$$tr\big(T(I)\big) = \sum \lambda_i = n, \qquad ds(T) = \sum (\lambda_i - 1)^2 > \epsilon$$

$$det\big(T(I)\big) = \prod \lambda_i \le exp(-\epsilon/6)$$

Analysis – Step 2

Claim: assume $T^*(I) = I$.

$$ds(T) > \epsilon \Rightarrow det\big(T(I)\big) \le exp(-\epsilon/6)$$

Step 2: assume $T^*(I) = I$ and $ds(T) > \epsilon$. Normalizing increases capacity by $exp(\epsilon/6)$.

Proof: Normalizing gives us

$$(A_1, \dots, A_m) \leftarrow (LA_1, \dots, LA_m), \qquad L = T(I)^{-1/2}$$

$$T_L(X) = \sum (LA_i) X (LA_i)^\dagger = T(I)^{-1/2} T(X) T(I)^{-1/2}$$

$$cap(T_L) = inf\left\{\frac{det\big(T_L(X)\big)}{det(X)} : \ X \succ 0\right\} = det\big(T(I)\big)^{-1} \cdot cap(T)$$

$$cap(T_L) \ge exp(\epsilon/6) \cdot cap(T)$$

Analysis – Step 3

Step 3: $T(X)$ normalized then $\mathbf{cap(T)} \leq \mathbf{1}$

$$T(X) = I \Rightarrow cap(T) = \inf_{X \succ 0} \frac{det\big(T(X)\big)}{det(X)} \leq \frac{det\big(T(I)\big)}{det(I)} = 1$$

$$T^*(X) = I \Rightarrow cap(T) \leq \frac{det\big(T(I)\big)}{det(I)} \leq \left(\frac{tr\big(T(I)\big)}{n}\right)^n$$

$$\left(\frac{tr\big(T(I)\big)}{n}\right)^n = \left(\frac{tr\big(T^*(I)\big)}{n}\right)^n = 1$$

$$tr\big(T(I)\big) = tr\big(\sum A_i A_i^\dagger\big) = tr\big(\sum A_i^\dagger A_i\big) = tr\big(T^*(I)\big) = n$$

Properties of Potential Function

Properties used by Potential Function:

1. **Zeroness/Nonzeroness gives answer to scaling**
2. **Invariant under $SL(n) \times SL(n)$:**
   - $\mathrm{cap}((L,R) \cdot T) = \det(L)^{2/n} \cdot \det(R)^{2/n} \cdot cap(T)$
3. **If nonzero, then far from zero:** if $cap(v) > 0$ for a vector with integer entries, then $cap(T) > exp(-poly(n))$

Which functions satisfy these conditions?

- **[Gurvits'04]** $cap(T) = inf\left\{\frac{det(T(X))}{det(X)} : X \succ 0\right\}$
- $cap'(T) = inf\{|| (L,R) \cdot T||^2 : (L,R) \in SL(n) \times SL(n)\}$

Are these functions the same? Yes!

Step 1 – Capacity Bounds From Invariants

Theorem 1: invariants of $\boldsymbol{SL(n)\times SL(n)}$ action generated by

$$p_Y(X_1,\ldots,X_m) = det\left(\sum_{1\le i\le m} Y_i\otimes X_i\right),\qquad Y_i\in M_d(\mathbb{C})$$

Theorem 2: $cap(T)>0\Leftrightarrow$ there exists $d\in\mathbb{N}$ and $Z_i\in M_d(\mathbb{C})$ s.t.

$$p_Z(A_1,\ldots,A_m) = det\left(\sum_{1\le i\le m} Z_i\otimes A_i\right)\neq 0$$

Lemma 3: can take $Z_i\in M_d(\mathbb{N})$ s.t. entries of Z_i are in $[dn+1]$.

$p_Z(X_1,\ldots,X_m)$ has integer coefficients!

Step 1 – Capacity Bounds From Invariants

So far: found invariant $p_Z(X_1, \dots, X_m)$ with integer coefficients and degree dn s.t.

$$|p_Z(A_1, \dots, A_m)| \geq 1$$

Claim: $p_Z(A_1, \dots, A_m) > 0 \Rightarrow ||C|| \geq exp(-poly(n))$ for any scaling $C = (C_1, \dots, C_m)$ of $(A_1, \dots, A_m)$.

Proof Sketch:

$$p_Z(C_1, \dots, C_m) = p_Z(A_1, \dots, A_m) \neq 0$$

$$\Rightarrow 1 \leq |p_Z(C_1, \dots, C_m)|$$

$$1 \leq |p_Z(C_1, \dots, C_m)| \leq \binom{n+d}{d} (dn)^{dn} \cdot ||C||^{dn}$$

$$\frac{1}{(dn)^{2dn}} \leq ||C||^{dn} \Rightarrow ||C|| \geq \frac{1}{(dn)^2}$$

Refining this analysis gives us no dependence on d.

Questions/Teasers

- Efficient algorithms for group actions (G, V) which are **not** product groups (like $SL(n) \times SL(n)$)?
  - Can use $cap(v) := inf\{||\ g \cdot v||^2 \ : g \in G\}$
  - Analog of DS?
  - See Michael's talk!

- Faster algorithms for scaling problems?
  - Algorithm shown has running time $poly(n \cdot 1/\epsilon)$
  - Can we get $poly(n \cdot log(1/\epsilon))$?
  - See afternoon talk!

Thank you!