Peter Bürgisser
Cole Franks
Ankit Garg
Leonid Gurvits
Rafael Oliveira
Michael Walter
Avi Wigderson

Alternating Minimization Algorithms for Matrix/Operator/Tensor Scaling & their Analysis

Contents

- Alternating Minimization
- Scaling Problems
- Scaling Algorithms
- Analysis
- Arbitrary Marginals

Alternating Minimization - Setup

Input: function

$$f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_d \to \mathbb{R}$$

Goal: find **global minimum** $(x_1^*, \dots, x_d^*)$ i.e.

$$f(x_1^*, \dots, x_d^*) \simeq \inf_{\substack{x_i \in \mathcal{X}_i \\ 1 \le i \le d}} f(x_1, \dots, x_d)$$

Minimizing f in **each block** is **simple**.

That is, for any $1 \le i \le d$, given $(y_1, \dots y_{i-1}, y_{i+1}, \dots y_d)$

easy to solve

$$\inf_{x_i \in \mathcal{X}_i} f(y_1, \dots, y_{i-1}, x_i, y_{i+1}, \dots, y_d)$$

Alternating Minimization - Heuristics

Repeatedly solve basic problem on different coordinates. Start with an initial guess $y^{(0)} = (y_1^{(0)}, \dots, y_d^{(0)})$ and time bound T

1. Given vector $y^{(t)} = \left(y_1^{(t)}, \dots, y_d^{(t)}\right)$, choose a coordinate $1 \leq i \leq d$
2. Find

$$z \simeq \arg\inf_{x_i \in \mathcal{X}_i} f(y_1^{(t)}, \dots, y_{i-1}^{(t)}, x_i, y_{i+1}^{(t)}, \dots, y_d^{(t)})$$

4. Set $y_i^{(t+1)} = z$, keep all other coordinates **unchanged**
5. If $t < T$ go back to step 1.

AM – product group actions

Setup: Group $G = G_1 \times G_2 \times \cdots \times G_d$ acting on vector space V (for instance $G_i = SL(n_i)$, $V = Ten(n_1, \ldots, n_d)$).

$$(A_1, \ldots, A_d) \cdot X \stackrel{\text{def}}{=} (A_1 \otimes \cdots \otimes A_d)X$$

Input: given $X \in V$, function

$$\mathrm{f_X}: G_1 \times \cdots \times G_d \to \mathbb{R}_{\geq 0}$$

$$\mathrm{f_X}(\mathrm{A_1}, \ldots, \mathrm{A_d}) = \|(\mathrm{A_1}, \ldots, A_d) \cdot X\|_2^2$$

Goal: find elt of min norm in $\overline{\mathcal{O}_G(X)}$, i.e., $(\mathrm{A}_1^*, \ldots, \mathrm{A}_\mathrm{d}^*)$ such that

$$\mathrm{f_X}\left(\mathrm{A}_1^*, \ldots, \mathrm{A}_\mathrm{d}^*\right) \simeq \inf_{\substack{A_i \in G_i \\ 1 \leq i \leq d}} f_X(A_1, \ldots, A_d) \stackrel{\text{def}}{=} cap(X)$$

Null-cone problem: $X \in \mathcal{N}_G(V) \Leftrightarrow cap(X) = 0$

KN'79 – Duality Theory

Capacity (primal): find elt of min norm in $\overline{\mathcal{O}_G(A)}$, i.e., $(A_1^*, \dots, A_d^*)$ such that

$$f_X(A_1^*, \dots, A_d^*) \simeq \inf_{\substack{A_i \in G_i \\ 1 \le i \le d}} f_X(A_1, \dots, A_d) \stackrel{\text{def}}{=} cap(X)$$

Moment map $\mu(X)$ at $X \in V$, define $h_X : G \to \mathbb{R}_{\ge 0}$ by

$$h_X(A_1, \dots, A_d) = \left\|\mu\big((A_1, \dots, A_d) \cdot X\big)\right\|_2^2$$

Moment map (dual): find elt in $\overline{\mathcal{O}_G(X)}$, i.e., $(A_1^*, \dots, A_d^*)$ that *minimizes norm* of moment map

$$h_X(A_1^*, \dots, A_d^*) \simeq \inf_{\substack{A_i \in G_i \\ 1 \le i \le d}} h_X(A_1, \dots, A_d) \stackrel{\text{def}}{=} cap_\mu(X)$$

[KN'79] $cap_\mu(X) = 0 \Leftrightarrow cap(X) > 0$

Non-Negative Matrices & Scaling

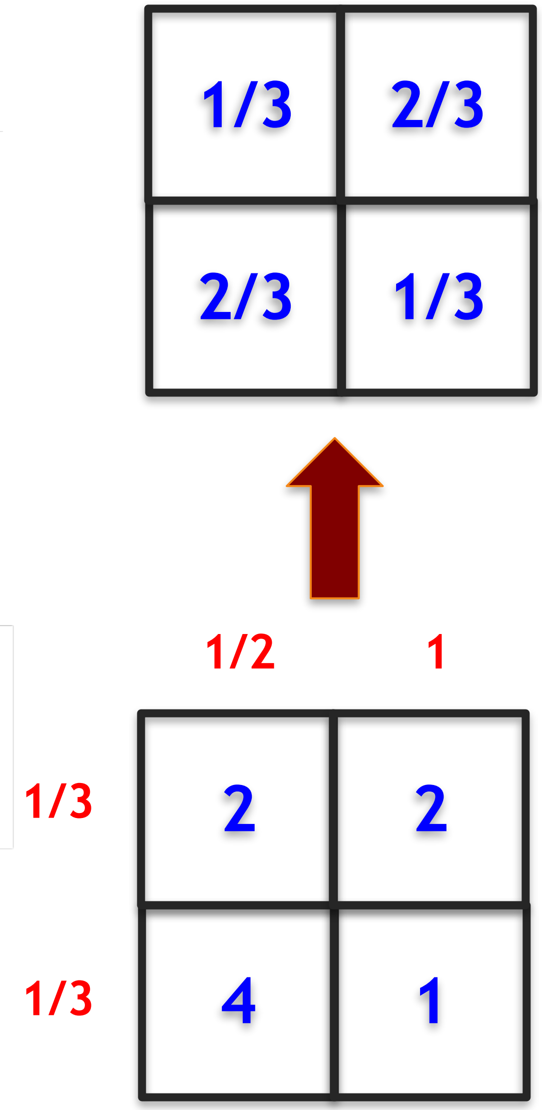
$X \in \mathrm{Mat}_n(\mathbb{R}_{\geq 0})$ is **doubly stochastic (DS)** if row/column sums of X are equal to **1/n**.

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

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

X has approx. DS scaling if $\forall \epsilon > 0$ there is scaling Y_ϵ of X s.t. all row/column sums of Y_ϵ are in $[1/n - \epsilon, 1/n + \epsilon]$.

1. When does X have approx. DS scaling?
2. Can we find it efficiently?



Matrix Scaling as null-cone problem

Group $G = ST(n) \times ST(n)$ acts on $V = Mat_n(\mathbb{C})$ by

$$\begin{pmatrix} \alpha_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \alpha_n \end{pmatrix} \cdot X \cdot \begin{pmatrix} \beta_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \beta_n \end{pmatrix}$$

Let $r_i = \frac{1}{\|X\|^2} \sum_j |x_{ij}|^2$ and $c_j = \frac{1}{\|X\|^2} \sum_i |x_{ij}|^2$ be X "row/column sums".

Moment Map (from Ankit's talk):

$$\boldsymbol{\mu}(\boldsymbol{X}) = \left(\boldsymbol{r}_1 - \frac{\mathbf{1}}{\boldsymbol{n}}, \cdots, \boldsymbol{r}_n - \frac{\mathbf{1}}{\boldsymbol{n}}, \boldsymbol{c}_1 - \frac{\mathbf{1}}{\boldsymbol{n}}, \cdots, \boldsymbol{c}_n - \frac{\mathbf{1}}{\boldsymbol{n}}\right)$$

Dual problem:

$$\boldsymbol{ds}(\boldsymbol{X}) = \|\boldsymbol{\mu}(\boldsymbol{X})\|^2 = \sum_i \left(\boldsymbol{r}_i - \frac{1}{\boldsymbol{n}}\right)^2 + \sum_j \left(\boldsymbol{c}_j - \frac{1}{\boldsymbol{n}}\right)^2$$

$\boldsymbol{X}$ has approx. DS scaling iff $\forall \epsilon > \mathbf{0}$, $\exists$ scaling $\boldsymbol{Y}_\epsilon$ s.t. $\mathbf{ds}(\mathbf{Y}_\epsilon) < \boldsymbol{\epsilon}$.

Matrix Scaling - Algorithm S

Problem: $X \in Mat_n(\mathbb{C})$, $\epsilon > 0$, is there ϵ-scaling to DS? If yes, find it.

Algorithm S [Sinkhorn'64]:

Repeat k times:

1. Normalize rows of X (make $r_i = 1/n$)
2. Normalize columns of X (make $c_j = 1/n$)

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

Else, output: **no scaling.**

Questions:

- Are we making progress at all?
- How do we know when to stop? (i.e., choose k)
- Is there an ϵ_0 such that if can scale to ϵ_0 then can scale for any ϵ?

Quantum Operators – Definition

A **quantum operator** is any map $\mathbf{T}: \mathbf{M}_n(\mathbb{C}) \to \boldsymbol{M}_n(\mathbb{C})$ given by $(\boldsymbol{A}_1, \dots, \boldsymbol{A}_m)$ s.t.

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

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

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

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

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

$$(\boldsymbol{A}_1, \dots, \boldsymbol{A}_m) \to (\boldsymbol{LA}_1\boldsymbol{R}, \dots, \boldsymbol{LA}_m\boldsymbol{R})$$

Operator Scaling

Moment Map (from Ankit's talk):

$$\mu(T) = (T(I_n) - \alpha I_n, T^*(I_n) - \alpha I_n), \alpha = tr(T(I_n))/n$$

Distance to doubly-stochastic:

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

$T(X)$ has approx. doubly stochastic scaling if

$$\inf_{T_{L,R}} ds(T_{L,R}) = 0$$

Once again, dual problem is the scaling problem!

1. When does $(A_1, \ldots, A_m)$ have approx. doubly stochastic scaling?
2. Can we find it efficiently?

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: $(A_1, \dots, A_m) \leftarrow (RA_1, \dots, RA_m)$ s.t. $T(I) = \alpha I$
2. Right normalize: $(A_1, \dots, A_m) \leftarrow (A_1C, \dots, A_mC)$ s.t. $T^*(I) = \alpha I$

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

Else output **no scaling**.

- Which k should we choose?
- Is there an ϵ_0 such that if can scale to ϵ_0 then can scale for any ϵ?

Tensor Scaling Problem

Let V $= Ten(n_1, \dots, n_d)$ and $G = SL(n_1) \times \cdots \times SL(n_d)$

$g = (A_1, \dots, A_d) \in \boldsymbol{G}$ acts on $X \in V$ in the natural way:

$$g \cdot X = (A_1, \dots, A_d) \cdot X \stackrel{\text{def}}{=} (A_1 \otimes \cdots \otimes A_d) X$$

Goal: given X, find $g^* = (A_1^*, \dots, A_d^*)$ such that

$$||g^* \cdot X||_2^2 \simeq cap(X) \stackrel{\text{def}}{=} \inf_{g \in G} ||g \cdot X||_2^2$$

Null-cone Problem: $cap(X) =^? 0$

Moment map (dual) Problem?

Quantum Setting (Previous talks)

Tensor $X \in Ten(n_1, \ldots, n_d)$ is **pure quantum state***, can be written as $\rho = XX^\dagger = |\psi\rangle\langle\psi|$.

- PSD matrix, dim $n = n_1 \cdots n_d$, $tr(\rho) = \|X\|^2$.

Let ρ_i be **marginal** of ρ with respect to particle i.

- PSD matrix, dim n_i, $tr(\rho_i) = \|X\|^2$.

ρ **d-stochastic (locally maximally entangled)** if all $\rho_i \propto I_{n_i}$

Quantum distillation: given pure state ρ, is there a scaling of ρ into a d-stochastic state?

Moment map (from Ankit's talk):

$$\mu(X) = \left(\frac{1}{\|X\|^2}\rho_1 - \left(\frac{1}{n_1}\right) I_{n_1}, \cdots, \frac{1}{\|X\|^2}\rho_d - \left(\frac{1}{n_d}\right) I_{n_d}\right)$$

Dual Problem: $ds(X) = \|\mu(X)\|^2 \rightarrow dds(X) = \inf_{Y \in \mathcal{O}(X)} ds(Y)$

Tensor Scaling – Algorithm

Problem: $\boldsymbol{X} \in \boldsymbol{Ten}(\boldsymbol{n_1}, \dots, \boldsymbol{n_d})(\mathbb{Z}[\boldsymbol{i}]), \boldsymbol{\epsilon} > \boldsymbol{0}$, is there ϵ-scaling to DS? If yes, find it.

Algorithm Q [BGOWW'18]:

Start with input X and scaling $(I_{n_1}, \dots, I_{n_d})$

Repeat $\boldsymbol{k}$ times:

1. If $\mathrm{ds}(\boldsymbol{X}) < \boldsymbol{\epsilon}$, output the scaling so far.
2. Let i be marginal s.t. $\left\|\frac{\mathbf{1}}{\|\boldsymbol{X}\|^2}\boldsymbol{\rho_i} - \frac{\mathbf{1}}{\boldsymbol{n_i}}\boldsymbol{I_{n_i}}\right\|^2 > \frac{\boldsymbol{\epsilon}}{\boldsymbol{n}}$
3. Normalize $\boldsymbol{\rho_i}$ (make $\rho_i = I_{n_i}$)

Output: **no scaling.**

Questions:

- How do we know when to stop? (i.e., choose k)
- Is there an ϵ_0 such that if can scale to ϵ_0 then can scale for any ϵ?

Analysis – General Approach

Three steps:

1. **[Upper bound]** in beginning $\|X\|^2 \leq poly(n, 2^b)$
   - Trivial from input data
2. **[Progress/step]** If $ds(X) > \epsilon$ (i.e., far from solution to dual) then normalization decreases $\|X\|^2$ by factor $\times \exp(O(\epsilon/n))$ (i.e., makes progress in primal)
   - Quantitative AM-GM (easy)
3. **[Lower bound]** $cap(X) > 0 \Rightarrow cap(X) > 1/n^2$
   - Invariant polynomials generated by nice poly. (hard)

ϵ-scaling problem $\rightarrow$ running time of $poly(nb/\epsilon)$.
Solve null-cone prb:
Matrix/Operator scaling: $\epsilon = O(1/n)$ is enough **[Gur'04]**
Tensor scaling: $\epsilon = \exp(-n \log n)$ **[HM, NM'84, BGOWW'18]**

Algorithm S – Analysis [LSW’00*]

Algorithm S [Sinkhorn’64]:

Repeat $\boldsymbol{k}$ times:

1. Normalize rows of $\boldsymbol{X}$ (make $r_i = 1/n$)
2. Normalize columns of $\boldsymbol{X}$ (make $c_j = 1/n$)

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

Else, output: **no scaling.**

Analysis [LSW’00*]:

1. $\|X\|^2 \leq \boldsymbol{poly}(\boldsymbol{n}, \boldsymbol{2^b})$ as $\boldsymbol{X}$ is integer bit comp. $\boldsymbol{b}$
2. $\mathbf{ds}(\boldsymbol{X}) > \boldsymbol{\epsilon} \Rightarrow \|\boldsymbol{X}\|^2$ **shrinks** by $\exp(\boldsymbol{O}(\boldsymbol{\epsilon}/\boldsymbol{n}))$ after normlztn
3. $\boldsymbol{Per}(\widehat{\boldsymbol{Y}}) \geq \mathbf{1}$ for any matrix $\boldsymbol{Y}$ in orbit of $\boldsymbol{X} \Rightarrow \|\boldsymbol{Y}\|^2 \geq \mathbf{1}/\boldsymbol{n}^2$

Within $\mathbf{k} = \boldsymbol{poly}(\boldsymbol{nb}/\boldsymbol{\epsilon})$ iterations we will get our scaling!

If $\text{Per}(\hat{X}) > 0 \Leftrightarrow X$ has no Hall blocker, so it is correct.

$\text{Per}(\hat{X})$ not needed, as any monomial encoding matching** works.

Algorithm G – Analysis [GGOW'16]

Algorithm G [Gurvits' 04]:

Repeat $\boldsymbol{k}$ times:

1. Left normalize: $(\boldsymbol{A_1}, \dots, \boldsymbol{A_m}) \leftarrow (\boldsymbol{RA_1}, \dots, \boldsymbol{RA_m})$ s.t. $\boldsymbol{T}(\boldsymbol{I}) = \boldsymbol{I}$.
2. Right normalize: $(\boldsymbol{A_1}, \dots, \boldsymbol{A_m}) \leftarrow (\boldsymbol{A_1 C}, \dots, \boldsymbol{A_m C})$ s.t. $\boldsymbol{T}^*(\boldsymbol{I}) = \boldsymbol{I}$.

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

Else output **no scaling**.

Analysis [GGOW'16]:

1. $\|\boldsymbol{T}\|^2 \leq \boldsymbol{poly}(\boldsymbol{n}, \boldsymbol{2^b})$ as $\boldsymbol{A_i}$ is integer bit comp. $\boldsymbol{b}$
2. $\mathbf{ds}(\boldsymbol{T}) > \boldsymbol{\epsilon} \Rightarrow \|\boldsymbol{X}\|^2$ **shrinks** by $\exp(\boldsymbol{O}(\boldsymbol{\epsilon}/\boldsymbol{n}))$ after normlztn
3. $\boldsymbol{P}(\boldsymbol{T_{L,R}}) \geq \mathbf{1}$ for any $\boldsymbol{T_{L,R}}$ in orbit of $\boldsymbol{T} \Rightarrow \|\boldsymbol{T_{L,R}}\|^2 \geq \mathbf{1}/\boldsymbol{n}^2$

Within $\mathbf{k} = \boldsymbol{poly}(\boldsymbol{nb}/\boldsymbol{\epsilon})$ iterations we will get our scaling!

- Is there an ϵ_0 such that if can scale to ϵ_0 then can scale for any ϵ?

Algorithm Q – Analysis [BGOWW'18]

Problem: $X \in Ten(n_1, \dots, n_d)(\mathbb{Z}[i]), \epsilon > 0$, is there ϵ-scaling to DS?
$n = n_1 n_2 \cdots n_d$.

Analysis:

1. $\|X\|^2 \le poly(n, 2^b)$ as X is integer bit comp. b
2. $ds(X) > \epsilon \Rightarrow \|X\|^2$ **shrinks** by $\exp(O(\epsilon/n))$ after normlztn
3. $P(Y) = P(X) \ge 1$ for any $Y \in \overline{\mathcal{O}(X)} \Rightarrow \|Y\|^2 \ge 1/n^2$

Within $k = poly(nb/\epsilon)$ iterations we will get our scaling!

Step 3:

Invariant ring $\mathbb{C}[X]^G$ generated by polys of:

1. degree $\le 2^{n^2}$ [Derksen'01]
2. Integer coefficients of norm $\le \mathrm{poly}(n)$ [Pro'07, BI'13,BGOWW'18]*

$$P(X) > 0, \deg(P) = m \Rightarrow P(X) \ge 1 \Rightarrow \|X\|^m \cdot n^{2m} \ge 1$$

Choosing ϵ for Null-cone problem

Problem: $\boldsymbol{X} \in \boldsymbol{Ten}(\boldsymbol{n_1}, \dots, \boldsymbol{n_d})(\mathbb{Z}[\boldsymbol{i}])$, is $\boldsymbol{X} \in \mathcal{N}_{\boldsymbol{G}}(\boldsymbol{V})$?

Algorithm Q [BGOWW'18]:

Start with input X and scaling $(I_{n_1}, \dots, I_{n_d})$

Repeat $\boldsymbol{k}$ times:

1. If $\mathrm{ds}(\boldsymbol{X}) < \boldsymbol{\epsilon}$, output the scaling so far.
2. Let i be marginal s.t. $\left\| \frac{\mathbf{1}}{\|X\|^2} \boldsymbol{\rho_i} - \frac{\mathbf{1}}{\boldsymbol{n_i}} \boldsymbol{I_{n_i}} \right\|^{\mathbf{2}} > \frac{\boldsymbol{\epsilon}}{\boldsymbol{n}}$
3. Normalize $\boldsymbol{\rho_i}$ (make $\rho_i = I_{n_i}$)

Output: **no scaling.**

Which ϵ should we choose?

1. **[Mum'65]** Instability parameter for tensor – "how quickly can we drive tensor to zero"
2. **[NM'84]** Instability lower bounds $\boldsymbol{ds}(\boldsymbol{X})$ for any $\boldsymbol{X}$
3. Bound on instability (bound soln to LP) implies $\boldsymbol{ds}(\boldsymbol{X})$ l.b.

Instability and ds(X) lower bound

[HM]: $\boldsymbol{X} \in \mathcal{N}_G(\boldsymbol{V}) \Leftrightarrow$ 1-PSG $\boldsymbol{\lambda} : \mathbb{C}^\times \to \boldsymbol{G}$ s.t. $\lim_{t\to 0} \boldsymbol{\lambda}(t)\, \boldsymbol{X} = \mathbf{0}$

[Mum'65]: how quickly does it go to zero?

$$\lambda(t) \leftarrow \left(B_i^{-1}\, diag(t^{a_{i1}}, \cdots, t^{a_{in}}) B_i\right)_{i=1}^{d},$$

$$B_i \in U(n_i), a_{ij} \in \mathbb{Z}, \sum_{j=1}^{n} a_{ij} = 0$$

1. $\mathrm{supp}((B_1, \cdots, B_d) \cdot X) \overset{\mathrm{def}}{=}$ set of nonzero entries
2. If for every $(j_1, \dots, j_d)$ in support, $\sum a_{ij_i} > 0$ then $\lambda(t) X \to \mathbf{0}$
3. Instability $\mathbf{inst}(\lambda, \mathbf{X}) = \min_{(j_1,\dots,j_d)\in supp} \frac{\sum a_{ij_i}}{\sqrt{\sum a_{ij}^2}}$

$$inst(X) = \max_{\lambda\ 1PSG} inst(\lambda, X)$$

Given 1-PSG $\boldsymbol{\lambda} : \mathbb{C}^\times \to \boldsymbol{G}$ s.t. $\lim_{t\to 0} \boldsymbol{\lambda}(t)\, \boldsymbol{X} = \mathbf{0}$ how to l.b. $\mathbf{inst}(\lambda, \mathbf{X})$

Instability and $\mathbf{ds(X)}$ lower bound

1-PSG $\boldsymbol{\lambda}$ s.t. $\lim_{t\to 0} \boldsymbol{\lambda}(\boldsymbol{t})\, \boldsymbol{X} = \mathbf{0} \Rightarrow$ following LP has soln

$$\boldsymbol{a_{ij}} \in \mathbb{Q}\,, \textstyle\sum_{j=1}^{n} \boldsymbol{a_{ij}} = \mathbf{0}\ \ \forall \boldsymbol{i} \in [\boldsymbol{d}],$$

$$\textstyle\sum_{i=1}^{d} \boldsymbol{a_{ij_i}} \geq \mathbf{1}\ \forall (\boldsymbol{j_1}, \dots, \boldsymbol{j_d}) \text{ in support}$$

Thus, has soln bounded by $\exp(-n \cdot \log(n))$ **[Sch'98]**

How does $\mathbf{inst(X)}$ l.b. $\mathbf{ds(X)}$?

Easy calculation shows that for any (a_{ij})

$$\min_{(j_1,\dots,j_d)\in supp(X)} \frac{\sum \boldsymbol{a_{ij_i}}}{\sqrt{\sum \boldsymbol{a_{ij}^2}}} \leq \sqrt{\boldsymbol{ds(X)}} = \sqrt{\boldsymbol{ds}\big((\boldsymbol{U_1}, \dots, \boldsymbol{U_d}) \cdot \boldsymbol{X}\big)}$$

[NM'84]: this holds more general group actions

Thus $\mathbf{ds(X)} \geq \mathbf{inst(X)^2} \geq \boldsymbol{exp(-2n \cdot log(n))}$.

Matrix and operator scaling

1-PSG $\boldsymbol{\lambda}$ s.t. $\lim_{t\to 0} \boldsymbol{\lambda}(\boldsymbol{t})\, \boldsymbol{X} = \mathbf{0} \Rightarrow$ following LP has soln

$$\boldsymbol{a_{ij}} \in \mathbb{Q}\,, \sum_{j=1}^{n} \boldsymbol{a_{ij}} = \mathbf{0}\ \ \forall \boldsymbol{i} \in [\boldsymbol{d}],$$

$$\sum_{i=1}^{d} \boldsymbol{a_{ij_i}} \geq \mathbf{1}\ \forall(\boldsymbol{j_1}, \dots, \boldsymbol{j_d}) \text{ in support}$$

Thus, has soln bounded by $\exp(-n \cdot \log(n))$ **[Sch'98]**

How can we better lower bound $\mathbf{inst(X)}$ in Matrix/Operator Scaling?

Construct a small solution (a_{ij}) to

$$\min_{(j_1,\dots,j_d)\in supp(X)} \frac{\sum \boldsymbol{a_{ij_i}}}{\sqrt{\sum \boldsymbol{a_{ij}^2}}}$$

Hint: use Hall blocker to find small solution.

Thus in these cases: $\mathbf{ds(X)} \geq \mathbf{inst(X)^2} \geq \mathbf{1}/\boldsymbol{n}^{\mathbf{2}}$.

Example - Matrix scaling

t^2	0	0
0	t^{-1}	0
0	0	t^{-1}

1	1	1
1	0	0
1	0	0

t^2	0	0
0	t^{-1}	0
0	0	t^{-1}

Common Hall blocker (shrunk subspace) in operator scaling yields same bound.

Non-uniform marginals [BFGOWW'18]

Tensor $X \in \mathbb{P}(Ten(n_1, \dots, n_d))$ is **pure quantum state**, can be written as $\rho = XX^\dagger / X^\dagger X = |\psi\rangle\langle\psi|$. Group $G = GL(n_1) \times \cdots \times GL(n_d)$

Quantum marginal problem: given pure state $\boldsymbol{X}$, target marginals $\boldsymbol{\rho_1}, \dots, \boldsymbol{\rho_d}$, (unit trace) is there a scaling of $\boldsymbol{X}$ with such marginals?

(Michael's talk): membership in moment/entanglement polytope of X

$$\boldsymbol{p_i} = \boldsymbol{spec}(\boldsymbol{\rho_i}), \boldsymbol{p} \in \boldsymbol{\Delta}(\boldsymbol{X})$$

Natural attempt: is there another representation where the moment maps for image of X $\in Ten(n_1, \dots, n_d)$ is given by $\boldsymbol{\rho_i^X} - \boldsymbol{p_i}$?

Shifting trick (from Michael's talk): $\boldsymbol{p} = (\boldsymbol{p_1}, \dots, \boldsymbol{p_d})$ spectrum of marginals, $p_i = \ell_i / k$ then Y $\stackrel{\text{def}}{=} X^{\otimes k} \otimes \boldsymbol{v}_{\ell^*}$ is such that

$$\mu(Y) = k \cdot \mu(X) + \mu(\boldsymbol{v}_{\ell^*}) = k\mu(X) + \ell^* = \left(\boldsymbol{\rho_i^X} - \boldsymbol{p_i}\right)_{i=1}^{d}$$

Reduces arbitrary marginal problem to uniform case!*

Non-uniform Tensor Scaling – Algorithm

Problem: $\boldsymbol{X} \in \boldsymbol{Ten}(\boldsymbol{n_1}, \dots, \boldsymbol{n_d})(\mathbb{Z}[\boldsymbol{i}]), \boldsymbol{\epsilon} > \mathbf{0}, \boldsymbol{p_1}, \dots, \boldsymbol{p_d}$ target spectra.
Is there ϵ-scaling to $\boldsymbol{p_1}, \dots, \boldsymbol{p_d}$? If yes, find it.

Algorithm Q+ [BFGOWW'18]:

Start with input X and random scaling $\mathrm{g} = (A_1, \dots, A_d)$

Repeat $\boldsymbol{T}$ times:

1. If $\mathrm{ds_p}(\boldsymbol{X}) < \boldsymbol{\epsilon}$, output the scaling so far.
2. Let i be marginal s.t. $\left\| \frac{\mathbf{1}}{\|X\|^2} \boldsymbol{\rho_i} - \boldsymbol{p_{i\uparrow}} \right\|^{\mathbf{2}} > \frac{\boldsymbol{\epsilon}}{\boldsymbol{n}}$
3. Normalize $\boldsymbol{\rho_i}$
   1. $\boldsymbol{\rho_i} \leftarrow \boldsymbol{p_{i\uparrow}^{-1/2}} \boldsymbol{R_i}$, where $\boldsymbol{R_i} \boldsymbol{R_i^{\dagger}} = \boldsymbol{\rho_i}$
   2. $\boldsymbol{R_i}$ upper triangular (Borel)

Output: **no scaling.**

Analysis – General Marginals

Three steps:

1. **[Upper bound]** in beginning $||g_0 \cdot X||^2 \leq poly(n, k, 2^b)$
   - Need g_0 for HWV not to vanish w.h.p. (Michael's talk)
     - Need degree bounds on HWV to show that g_0 is nice
2. **[Progress/step]** If $ds(Y) > \epsilon$ (i.e., far from solution to dual) then normalization decreases $||Y||^2$ by factor $\times \exp(O(\epsilon/n))$ (i.e., makes progress in primal)
   - Quantitative AM-GM (easy)
   - Scale by Borel to keep HWV invariant under scaling
3. **[Lower bound]** $cap_p(X) > 0 \Rightarrow cap_p(X) > 1/n^2$
     - HWVs are generated by nice poly. (hard)

Thank you!