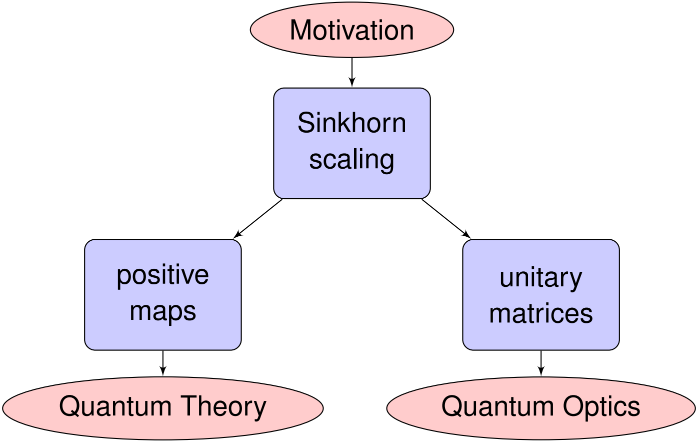
Sinkhorn's Theorem in Quantum Information Theory

Martin Idel

TNG Technology Consulting (work done at TUM)

21 March 2018

What to expect?



How can we get the real matrix from its marginals?

From/To	England	Wales	Scotland	Total
England				100,000
Wales				50,000
Scotland				50,000
Total	50,000	75,000	75,000	

How can we get the real matrix from its marginals?

From/To	England	Wales	Scotland	Total
England				100,000
Wales				50,000
Scotland				50,000
Total	50,000	75,000	75,000	

From/To	England	Wales	Scotland	Total
England	500	70	300	870
Wales	20	200	100	320
Scotland	10	30	50	90
Total	530	300	450	

What's the best guess to fill the table?

There is a very simple algorithm

Algorithm: Given A

- Noramlise column sums (multiply column j by $1/\sum_i A_{ij}$ from the left)
- Normalise row sums (multiply row i by $1/\sum_j A_{ij}$ from the right)

Matrix estimation

Problem

Let $A \in \mathbb{R}^{n \times n}$ be a positive matrix. Do there exist positive diagonal matrices D_1, D_2 such that $D_1 A D_2$ is doubly stochastic?

- transportation matrices [R. Kruithof, *De Ingenieur*, 52, 1937]
- voting systems
- conditioning matrices
- Markov chain estimation [R. Sinkhorn, *Annals of Mathematical Statistics*, 35, 1964]

Schrödinger bridges

- discretised, finite phase space
- a priori “guess” of (one-step) stochastic process (Brownian motion)
- the densities at t_0 and t_1 distributions do not fit

Problem:

What probability distribution fits best to the observed data and the a priori distribution? Use

$$D(p||q) = \sum_i p_i \ln\left(\frac{p_i}{q_i}\right)$$

as measure for “best fit” (large deviations).

Proposed by Schrödinger [*Sonderausgabe a. d. Sitz.-Ber. d. Preuß. Akad. d. Wiss., Phys.-math. Klasse*, 1931]

Schrödinger bridges

- discretised, finite phase space
- a priori "guess" of (one-step) stochastic process (Brownian motion)
- the densities at t_0 and t_1 distributions do not fit

Problem:

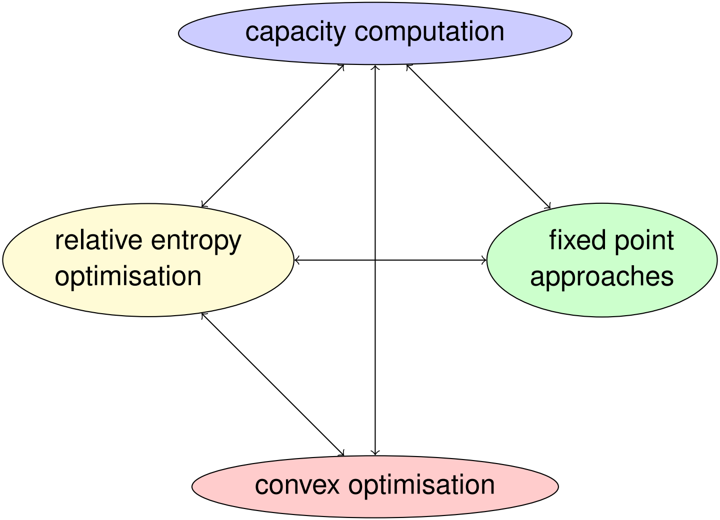
What probability distribution fits best to the observed data and the a priori distribution? Use

$$D(p||q) = \sum_i p_i \ln\left(\frac{p_i}{q_i}\right)$$

as measure for "best fit" (large deviations).

Proposed by Schrödinger [*Sonderausgabe a. d. Sitz.-Ber. d. Preuß. Akad. d. Wiss., Phys.-math. Klasse*, 1931]
Answer: dual to matrix scaling [T. Georgiou, M. Pavon, *JMP*, 56, 2015].

There are various ways to approach matrix scaling



Scaling boils down to a simple fixed point theorem

Lemma ([M.V. Menon, *Proc. AMS*, 18, 1967])

Let $A \in \mathbb{R}^{n\times n}$ be a nonnegative matrix. Then there exists a scaling to a doubly-stochastic matrix iff there exist vectors x, y such that

$$Ax = y^{-1} \quad A^{tr}y = x^{-1} \qquad x_i > 0, y_i > 0 \;\forall i$$

Results can be phrased in terms of zero-patterns

Theorem ([R. Brualdi, *Journal of Mathematical Analysis and App.*, 16, 1966])

Let $A \in \mathbb{R}^{n\times n}$ be a nonnegative matrix. Then there exists a scaling to a doubly-stochastic map iff A has a doubly stochastic zero-pattern.

Theorem ([M.V. Menon, H. Schneider, *LAA*, 2, 1969])

Let $A \in \mathbb{R}^{n\times n}$ be a nonnegative matrix. There exist diagonal matrices D_1, D_2 such that $D_1 A D_2$ has marginals $p, q \in \mathbb{R}^n$ iff there exists a nonnegative matrix $B \in \mathbb{R}^{n\times n}$ with marginals p, q and the same zero-pattern as A.

Some remarks about the algorithm

- The algorithm converges and works perfectly fine.
- There exist more sophisticated algorithms with near linear time ($\mathcal{O}(m + n^{4/3})$)
- The decision problem of scalability is NP-hard. Approximate scalability is in P.

Extending doubly stochastic matrices to positive maps

Definition

Let $\mathcal{E} : \mathcal{M}_n \to \mathcal{M}_n$ be a linear map.

- $\mathcal{E}$ is called *positive*, if for all $A \geq 0$, $\mathcal{E}(A) \geq 0$.
- $\mathcal{E}$ is called *positivity improving*, if for all $A \geq 0$, $\mathcal{E}(A) > 0$ (i.e. full rank).
- $\mathcal{E}$ is called *fully indecomposable*, if it is positive and for all $A \in \mathcal{M}_n$ with $A \geq 0$ and $0 < \mathrm{rank}(A) < n$, we have $\mathrm{rank}(\mathcal{E}(A)) > \mathrm{rank}(A)$.

Note that these comprise *completely positive maps*, i.e. maps such that $\mathrm{id}_n \otimes \mathcal{E}$ is still positive for any $n \in \mathbb{N}$.

Extending doubly stochastic matrices to positive maps

Definition

Let $\mathcal{E} : \mathcal{M}_n \to \mathcal{M}_n$ be a linear map.

- $\mathcal{E}$ is called *positive*, if for all $A \geq 0$, $\mathcal{E}(A) \geq 0$.
- $\mathcal{E}$ is called *positivity improving*, if for all $A \geq 0$, $\mathcal{E}(A) > 0$ (i.e. full rank).
- $\mathcal{E}$ is called *fully indecomposable*, if it is positive and for all $A \in \mathcal{M}_n$ with $A \geq 0$ and $0 < \operatorname{rank}(A) < n$, we have $\operatorname{rank}(\mathcal{E}(A)) > \operatorname{rank}(A)$.

Note that these comprise *completely positive maps*, i.e. maps such that $\mathrm{id}_n \otimes \mathcal{E}$ is still positive for any $n \in \mathbb{N}$.

Question:

Given a positive linear map $\mathcal{E} : \mathcal{M}_n \to \mathcal{M}_n$, do there exist matrices $X, Y \in \mathcal{M}_n$ such that $\mathcal{E}'(\cdot) := Y^\dagger \mathcal{E}(X \cdot X^\dagger) Y$ is doubly stochastic ($\mathcal{E}'(\mathbb{1}) = \mathbb{1}$ and $\mathcal{E}'^*(\mathbb{1}) = \mathbb{1}$)?

Theorems

Theorem ([L. Gurvits, *J. of Comp. and Sys. Sci.*, 2004])

Let $\mathcal{E} : \mathcal{M}_n \to \mathcal{M}_n$ be a positive, linear map. Then there exists a unique scaling (up to constant and unitaries) to a doubly stochastic map iff $\mathcal{E}$ is fully indecomposable.

Partial results are also present in [M.I., *Master's thesis*, 2013]]

Theorems

Theorem ([L. Gurvits, *J. of Comp. and Sys. Sci.*, 2004])

Let $\mathcal{E}: \mathcal{M}_n \to \mathcal{M}_n$ be a positive, linear map. Then there exists a unique scaling (up to constant and unitaries) to a doubly stochastic map iff $\mathcal{E}$ is fully indecomposable.

Partial results are also present in [M.I., *Master's thesis*, 2013]]

Theorem

Let $\mathcal{E}: \mathcal{M}_n \to \mathcal{M}_n$ be a positivity improving, linear map. Given $V, W > 0$ with $\mathrm{tr}(V) = \mathrm{tr}(W)$, there exist scalings to a positive map $\mathcal{T}$ such that $\mathcal{T}(V) = W$ and $\mathcal{T}^*(\mathbb{1}) = \mathbb{1}$.

Implicit in [T. Georgiou, M. Pavon *JMP*, 56, 2015].

We can give a natural algorithm: $\mathcal{E}$ positive, linear map:

$$\begin{aligned}
\mathcal{E}_0 &:= \mathcal{E} \\
\mathcal{E}_{2i+1}(\cdot) &:= (\mathcal{E}_{2i}(\mathbb{1}))^{-1/2}\mathcal{E}_{2i}(\cdot)(\mathcal{E}_{2i}(\mathbb{1}))^{-1/2} \quad \forall i \geq 0 \\
\mathcal{E}^*_{2i}(\cdot) &:= (\mathcal{E}^*_{2i-1}(\mathbb{1}))^{-1/2}\mathcal{E}^*_{2i-1}(\cdot)(\mathcal{E}^*_{2i-1}(\mathbb{1}))^{-1/2} \quad \forall i \geq 1
\end{aligned}$$

We can give a natural algorithm: $\mathcal{E}$ positive, linear map:

$$\begin{aligned}
\mathcal{E}_0 &:= \mathcal{E} \\
\mathcal{E}_{2i+1}(\cdot) &:= (\mathcal{E}_{2i}(\mathbb{1}))^{-1/2}\mathcal{E}_{2i}(\cdot)(\mathcal{E}_{2i}(\mathbb{1}))^{-1/2} \quad \forall i \geq 0 \\
\mathcal{E}^*_{2i}(\cdot) &:= (\mathcal{E}^*_{2i-1}(\mathbb{1}))^{-1/2}\mathcal{E}^*_{2i-1}(\cdot)(\mathcal{E}^*_{2i-1}(\mathbb{1}))^{-1/2} \quad \forall i \geq 1
\end{aligned}$$

Theorem ([L. Gurvits, *J. of Comp. and Sys. Sci.*, 2004])

This algorithm converges to a unique doubly-stochastic map iff $\mathcal{E}$ is fully-indecomposable.

Proposition ([M.I., *arXiv:1609.06349*, 2016])

If $\mathcal{E}$ is positivity improving, we can say even more:

- *The algorithm converges geometrically in Hilbert's projective metric d_H, i.e.*

$$d_H(\mathcal{E}_k(\mathbb{1}),\mathbb{1}) + d_H(\mathcal{E}_k^*(\mathbb{1}),\mathbb{1}) \leq \frac{\gamma^k}{1-\gamma}(d_H(\mathcal{E}(\mathbb{1}),\mathbb{1}) + d_H(\mathcal{E}^*(\mathbb{1}),\mathbb{1}))$$

  with the contraction ratio $\gamma := \sup\{d_H(\mathcal{E}(X),\mathcal{E}(Y))/d_H(X,Y)\} < 1$.
- *The resulting doubly stochastic map is continuous in $\mathcal{E}$.*
- *For $k \to \infty$, $\mathcal{T}^k(X)/\operatorname{tr}(\mathcal{T}^k(X))$ converges to the unique eigenvector of $\mathcal{T}(\rho) := \mathcal{E}^*(\mathcal{E}(\rho)^{-1})^{-1}$.*

All results are generalisations of known matrix results (see e.g. [B. Lemmens, R. Nussbaum Nonlinear Perron-Frobenius Theory, 2012]).

Choi matrices

Using the *Choi-Jamiolkowski isomorphism*, one can get a normal form for block-positive matrices (extending [A. Kent et al. *PRL*, 1999]): There exist X_1, X_2 such that for every block-positive ρ with unit trace, we have:

$$(X_1 \otimes X_2)\rho(X_1 \otimes X_2)^\dagger = \frac{1}{d^2}\mathbb{1} + \sum_{k=1}^{k^2-1} \xi_k J_k^1 \otimes J_k^2$$

with traceless J_k.

Edmond's problem

Under certain assumptions, given a linear subspace $V := \langle V_1, \ldots, V_n \rangle \subset \mathcal{M}_n$ one can efficiently decide, whether the subspace contains a full-rank matrix or not.

More precisely:

Definition ([A. Widgerson, *Tutorial at CCC'17*, 2017])

Let $\mathbb{F}$ be a field and $\mathbf{x} = \{x_1, \ldots, x_m\}$ be as set of variables. Let $(A_1, \ldots, A_m)$ be a tuple of elements in $\mathcal{M}_n(\mathbb{F})$ and define

$$L = A_1 x_1 + \ldots A_m x_m$$

then Edmond's problem is the question whether $\det(L) \equiv 0$ over the field of polynomials $\mathbb{F}[\mathbf{x}]$.

Furthermore, we can define a noncommutative version of this question by asking whether the operator is invertible over the free skew field over $\mathbf{x}$.

There is a close connection to scaling

Theorem

Given a bipartite $n \times n$ graph with Boolean matrix A_G and symbolic matrix L_G such that the entry is x_{ij} if G has edge (i,j) and 0 otherwise, then

$$G \text{ has a perfect matching} \quad \Leftrightarrow \quad \det(L_G) \not\equiv 0 \text{ over } \mathbb{F}[\boldsymbol{x}] \quad \Leftrightarrow \quad A_G \text{ is scalable}$$

Theorem ([A. Garg et al. *FOCS*, 2016])

Given a noncommutative operator $L = A_1 x_1 + \ldots A_m x_m$, then it is not invertible over the free skew field iff the operator $\tilde{L}(X) := \sum_i A_i X A_i^\dagger$ is approximately scalable.

Questions I'd like to have answered

- What is the correct generalisation of zero-patterns to positive maps?
- Is there a way to view operator scaling as entropy problem?
- Can we set up “nonlinear noncommutative Perron-Frobenius theory”?
- Can we extend the fixed point map to the boundary?

Sinkhorn for unitary matrices

Definition

Let $U \in U(n)$ be a unitary matrix. We call it *quasi doubly-stochastic* if every row and column sum is one.

Theorem ([M.I., M.M. Wolf, *LAA*, 471, 2015])

Let $U \in U(n)$, then there exist diagonal unitaries L, R such that LUR is quasi doubly-stochastic.

Proposed in [A.D. Vos, S.D. Baerdemacker, *Open System & Information*, 2014].

It's similar to a fixed point problem

Lemma

Let $U \in U(n)$ and $T^n := \{\varphi \in \mathbb{C}^n | \, |\varphi_i| = 1, \forall i\}$, then there exists a scaling of U to a quasi doubly-stochastic matrix iff there exists $\varphi \in T^n$ such that $U\varphi \in T^n$

Symplectic ingredients

Definition

- A symplectic manifold $(\mathcal{M}, \omega)$ is a differentiable manifold with a closed, nondegenerate 2-form ω, e.g.

$$\omega = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

- A map $\psi : \mathcal{M} \to \mathcal{M}$ of a symplectic manifold is called a symplectomorphism, if $\psi^* \omega = \omega$.
- If $(\mathcal{M}, \omega)$ is a symplectic manifold of dimension $2n$, a submanifold $\mathcal{L}$ of dimension n is called *Lagrangian*, if $\omega_L = 0$.

Symplectic Geometry talks a lot about fixed points

Conjecture (Arnold's conjecture)

A symplectomorphism that is generated by a time-dependent Hamiltonian vector field should have at least as many fixed points as a function must have critical points

and for submanifolds:

Conjecture (Arnold's conjecture)

Let $L_0, L_1 \subset M$ be compact Lagrangian submanifolds of a symplectic manifold (M, ω). Suppose that ψ is a symplectomorphism corresponding to a compactly supported Hamiltonian which moves L_0 onto L_1, then L_0 and L_1 must have at least as many intersection points as a function on L_0 must have critical points.

The Clifford-Torus is a Lagrangian submanifold for which the conjecture holds

Theorem ([P. Biran et al. *Comm. in Contemp. Math.*, 6, 2004])

Let $\mathbb{C}P^n$ be the complex projective space with its torus T^n and its usual symplectic structure, then

$$T^n \cap \psi T^n \neq \emptyset$$

for any symplectomorphism ψ in the connected component of the identity in the group of symplectomorphisms.

The decomposition can be exploited recursively

Corollary

Let $U \in U(n)$, then there exist $C_1, C_2 \in \mathrm{Circ}(n)$ and $\tilde{U} \in U(n-1)$ such that

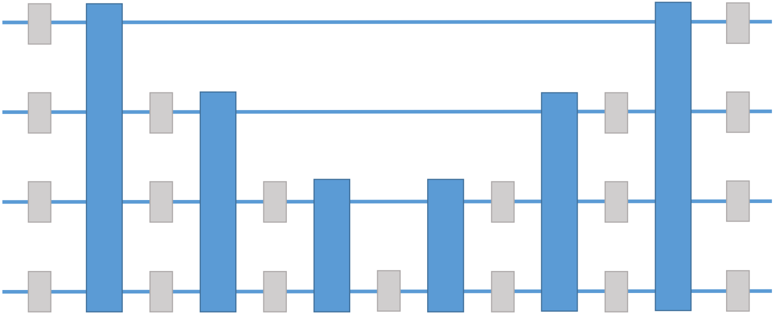
$$U = C_1 \operatorname{diag}(1, \tilde{U}) C_2. \tag{1}$$

Corollary

Let $U \in U(n)$, then there exist diagonal matrices $L, R \in U(n)$ and $\tilde{U} \in U(n-1)$ such that

$$U = L F_n \operatorname{diag}(1, \tilde{U}) F_n^{\dagger} R. \tag{2}$$

Application in Quantum Optics



Its place within other decomposition

Drawbacks:

- Not natural in qubit systems (does not decompose into tensor products)
- Requires multi-gate operators

It's a new normal form

- It seems to be unrelated to Lie theory (although the maximal torus pops up)
- It seems to be unrelated to canonical angles and the cosine-sine decomposition
- It seems to be unrelated to beam splitter transformations

When nomalising properly, we get a Sinkhorn-like algorithm

- Take row sums y_i and normalise $\Phi(y_i) = \exp(i \arg(y))$.
- Multiply from the right with normalised row sums on the diagonal.
- Repeat with column sums from the left

Potential approach: $\mathrm{sum}(X) = \sum_{i,j=1}^n X_{ij}$, then this is strictly increasing under this algorithm ([A.D. Vos, S.D. Baerdemacker *Open System & Information*, 2014]).
Unknown: Does it converge to n^2, which is the sum for matrices with row and column sums $= 1$.

There is another name in harmonic analysis

Definition

A **biunimodular vector** is a unimodular vector (i.e. all entries have modulus one) such that its Fourier transform is also a unimodular vector

Other names:

- CAZAC (constant amplitude zero autocorrelation) sequences,
- PSK (phase-shift keyed) perfect sequences,
- polyphase sequences with optimum correlation
- ...

Famous examples: Gauss sequences in dimension n

$$u_k = c \exp\left(\frac{2\pi i}{n}(\lambda k^2 + \mu k)\right) \qquad n \text{ odd} \tag{3}$$

$$u_k = c \exp\left(\frac{2\pi i}{n}(\lambda/2k^2 + \mu k)\right) \qquad n \text{ odd} \tag{4}$$

Extensions to other matrices yield Sinkhorn's problem for unitary matrices

[H. Führ, Rzeszotnik *LAA*, 484 2015].

The decompositions are not unique

- If all torus intersections are transversal, the lower bound is 2^{n-1} [C.H. Cho *Int. Math. Res. Not.*, 2004].
- For almost all $A \in U(n)$, the number of decompositions is finite (via Sard's theorem).
- Real parameter families of decompositions known if n is divisible by a square.
- We know some numbers for Fourier decompositions: For instance for F_7, it is 532, for F_{13}, the number is known to be 53222.

Conjecture (adapted from [G. Björck, B. Saffari, *C. R. Acad. Sci.*, 1995])

If $n \in \mathbb{N}$ is not divisible by a square, then the set of Sinkhorn decompositions up to a scalar multiple is finite.

There is a faint connection to MUBs

"Mutually unbiased basis" are sets of vectors $\{|a_n\rangle\}$, $\{|b_n\rangle\}$ which fulfil

$$|\langle a_i|b_j\rangle|^2 = 1/d, \quad \forall n$$

There is a faint connection to MUBs

"Mutually unbiased basis" are sets of vectors $\{|a_n\rangle\}$, $\{|b_n\rangle\}$ which fulfil

$$|\langle a_i|b_j\rangle|^2 = 1/d, \quad \forall n$$

We have:

Proposition ([K. Korzekwa et al. PRA 5, 2014])

For any two bases $\{|a_n\rangle\}$, $\{|b_n\rangle\}$ of a Hilbert space with dimension d, there exist states $|\psi\rangle$ that are unbiased with respect to both of them, i.e.

$$|\langle a_n|\psi\rangle|^2 = |\langle b_n|\psi\rangle|^2 = 1/d, \quad \forall n.$$

We need to add some more structure to actually say something meaningful about the number of mutually unbiased bases.

Questions I'd like to have answered

- What type of decomposition is it?
- How to prove this with linear algebra tools?
- Can any of the arguments from harmonic analysis be used?
- Does the algorithm converge?